

**Univerzita Karlova**

**Filozofická fakulta**

Ústav informačních studií a knihovnictví

Studijní program: Informační studia a knihovnictví

Studijní obor: Informační věda

**PhDr. Mgr. Ladislav Cubr**

**AUTENTICITA A DIGITÁLNÍ INFORMACE**  
**AUTHENTICITY AND DIGITAL INFORMATION**

Disertační práce

Školitel: prof. RNDr. Jiří Ivánek, CSc.

Praha 2017



Prohlášení:

Prohlašuji, že jsem disertační práci napsal samostatně s využitím pouze uvedených a řádně citovaných pramenů a literatury a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne

.....

Jméno a příjmení



## **Identifikační záznam**

CUBR, Ladislav. *Autenticita a digitální informace [Authenticity and digital information]*. Praha, 2017. 227 s. Disertační práce. Univerzita Karlova, Filozofická fakulta, Ústav informačních studií a knihovnictví. Školitel prof. RNDr. Jiří Ivánek, CSc.

### **Abstrakt:**

Disertace je zaměřena na problematiku autenticity digitalizátů knih, a to v celém jejich životním cyklu (produkce, archivace, zpřístupňování). Nejprve je přiblížen základní konceptuální rámec pro řízení životního cyklu digitálních dokumentů v organizacích (norma ISO 14721). Dále je představen současný stav problematiky řízení životního cyklu digitalizátů knih (tedy předmět autenticity). Poté jsou představeny relevantní tematizace autenticity digitálních dokumentů a provedena jejich analýza a revize. Výsledky této revize (obecný rámec pro zkoumání autenticity) jsou uplatněny pro specifikaci požadavků autenticity pro digitalizáty knih. Tato konceptualizace zahrnuje podrobnou analýzu problémů, které ohrožují autenticitu v životním cyklu digitalizovaných dokumentů. Na základě vybraných zjištění této konceptualizace je následně vypracován doporučený postup pro udržení autenticity digitalizátů knih, obsahující obecné principy i doporučení specifická pro konkrétní fáze jejich životního cyklu. Tento postup je dále podrobněji specifikován pro jedno z dílčích řešení problému zachování autenticity digitalizovaných dokumentů v celém jejich životním cyklu, kterým je systém perzistentní identifikace.

### **Abstract:**

The dissertation focuses on the authenticity of digitized books in the context of their life cycle (production, preservation, access). First the OIAS high-level conceptual framework for lifecycle management of digital documents maintained by organizations is introduced. Then the current situation of digitized books lifecycle management is described. This part is followed by an introducing to relevant conceptualizations of the authenticity of digital documents and these conceptualizations are analyzed and reviewed. Then framework for analysis of authenticity is established based on previous findings. This framework is then used to identify authenticity requirements for digitized books and to develop a domain-specific conceptualization of the authenticity of digitized books. This conceptualization deploys detailed analysis of risks threatening authenticity during lifecycle management of digitized books. The selected topics of this conceptualization are then the source for the next step,

which is to develop a recommended practice for maintaining authenticity of digitized books. This practice is further specified for one partial solution to the problem of maintaining the authenticity of digital documents throughout their life cycle, which is a persistent identification system.

**Klíčová slova:**

Autenticita, dlouhodobé uchovávání digitálních dokumentů, digitální archivace, digitalizáty knih, perzistentní identifikátory, model OAIS, norma ISO 14721.

**Keywords:**

Authenticity, long-term digital preservation, digital archiving, digitized books, persistent identifiers, OAIS model, ISO 14721 standard.



# OBSAH

<b>PŘEDMLUVA.....</b>	<b>1</b>
<b>1 ÚVOD .....</b>	<b>5</b>
<b>2 ZÁKLADNÍ KONCEPTUÁLNÍ RÁMEC ŽIVOTNÍHO CYKLU .....</b>	<b>9</b>
2.1 Historie norem .....	9
2.2 Norma ISO 14721 .....	11
2.2.1 Úvodní přehled .....	11
2.2.2 Informační model .....	19
2.2.3 Funkční model .....	26
2.2.4 Přístupy k uchovávání .....	34
2.3 Norma ISO 16363 .....	37
<b>3 SLOŽKY DIGITALIZÁTŮ KNIH V SOUČASNÉ PRAXI.....</b>	<b>40</b>
3.1 Formáty.....	41
3.1.1 Roviny a aspekty užití formátu .....	42
3.1.2 Výběr archivačního formátu .....	44
3.1.3 Doporučované a užívané archivační formáty .....	46
3.1.4 Prezentační formáty .....	47
3.1.5 Formátové registry .....	48
3.2 Obrazová data .....	49
3.2.1 Generace obrazových dat .....	50
3.2.2 Strukturální model obrazové reprezentace .....	50
3.2.3 Formátový profil .....	51
3.2.4 Typy komprese.....	51
3.2.5 Obecné obrazové vlastnosti.....	52
3.2.6 Prezentační varianty .....	54
3.3 Textová data jako výstup OCR.....	55
3.4 Metadata .....	56
3.4.1 Užívané metadatové standardy .....	57
3.4.2 Metadatový aplikační profil .....	57
3.4.3 PREMIS.....	58
3.4.4 METS.....	60
3.4.5 MIX.....	60
3.5 Metadata v obrazových souborech .....	61
<b>4 ETAPY ŽIVOTNÍHO CYKLU DIGITALIZÁTŮ KNIH .....</b>	<b>62</b>
4.1 Produkce .....	63
4.1.1 Základní aspekty produkce .....	63
4.1.2 Fáze snímání .....	63
4.1.3 Fáze zpracování obrazových dat .....	64
4.1.4 Fáze vytváření metadat .....	67



4.1.5	Fáze vytváření balíčku SIP a jeho dodávání do archivu.....	68
<b>4.2</b>	<b>Archivace .....</b>	<b>69</b>
4.2.1	Technické řešení archivace .....	69
4.2.2	Opatření vykonávaná na balíčku SIP v archivu .....	70
4.2.3	Opatření vykonávaná v průběhu uchovávání .....	70
4.2.4	Metadata pro archivaci.....	71
<b>4.3</b>	<b>Zpřístupňování .....</b>	<b>71</b>
4.3.1	Způsoby zpřístupňování v digitální knihovně.....	72
4.3.2	Metadata v digitální knihovně .....	73
<b>5</b>	<b>KONCEPT AUTENTICITY DIGITÁLNÍCH DOKUMENTŮ .....</b>	<b>74</b>
<b>5.1</b>	<b>Historie pojetí autenticity dokumentu .....</b>	<b>74</b>
<b>5.2</b>	<b>Autenticita v oblasti archivnictví .....</b>	<b>75</b>
5.2.1	Projekt Preservation of the Integrity of Electronic Records .....	76
5.2.2	Projekt InterPARES.....	79
<b>5.3</b>	<b>Autenticita v normě ISO 14721.....</b>	<b>88</b>
<b>5.4</b>	<b>Autenticita ve standardu PREMIS.....</b>	<b>91</b>
<b>5.5</b>	<b>Autenticita v knihovnictví.....</b>	<b>95</b>
5.5.1	Autenticita a uchovávání .....	95
5.5.2	Autenticita a zpřístupnění.....	98
5.5.3	Autenticita a digitalizace.....	99
5.5.4	Autenticita původních snímků .....	101
<b>5.6</b>	<b>Další vybrané tematizace autenticity.....</b>	<b>102</b>
5.6.1	Benjaminova koncepce jedinečného originálu .....	102
5.6.2	Autenticita ve směrnici DSA.....	103
5.6.3	Autenticita zdrojů výzkumného procesu .....	104
5.6.4	Rothenberg a autenticita spojená s vhodností .....	106
5.6.5	Gladney a struktura tvrzení o autenticitě .....	108
<b>5.7</b>	<b>Shrnutí a dílčí závěr .....</b>	<b>109</b>
<b>6</b>	<b>REVIZE KONCEPTŮ AUTENTICITY .....</b>	<b>112</b>
<b>6.1</b>	<b>Kontext pro zkoumání autenticity .....</b>	<b>112</b>
6.1.1	Aspekt institucionalizace .....	112
6.1.2	Knihy jako dokument .....	113
<b>6.2</b>	<b>Revize tematizací autenticity.....</b>	<b>115</b>
6.2.1	Model pro předmět autenticity .....	115
6.2.2	Časoprostorový aspekt .....	117
6.2.3	Otázka originálu .....	118
6.2.4	Aspekt odvozeniny.....	121
6.2.5	Koncept věrnosti digitalizace .....	122
6.2.6	Klíčové vlastnosti .....	124
6.2.7	Identita dokumentu .....	125
6.2.8	Integrita dokumentu.....	126
6.2.9	Koncept provenience .....	127
6.2.10	Dokumentace a metadata.....	128
6.2.11	Problémy reprodukce .....	129

6.2.12	Rizika pro udržení autenticity .....	130
6.2.13	Shrnutí revize a dílčí závěr .....	132
<b>6.3</b>	<b>Návrh obecného rámce pro zkoumání autenticity digitalizátů knih .....</b>	<b>134</b>
6.3.1	Model pro předmět autenticity v celém životním cyklu .....	134
6.3.2	Referenční objekt .....	135
6.3.3	Složky autenticity .....	136
6.3.4	Archivační a interpretační informace .....	137
<b>7</b>	<b>SPECIFIKACE POŽADAVKŮ AUTENTICITY PRO DIGITALIZÁTY KNIH .....</b>	<b>138</b>
<b>7.1</b>	<b>Stanovení referenčního objektu .....</b>	<b>138</b>
7.1.1	Úroveň vyjádření .....	139
7.1.2	Úroveň manifestace (vydání) a exempláře .....	140
7.1.3	Stanovení klíčových vlastností .....	141
<b>7.2</b>	<b>Identita digitalizátu knihy jako intelektuální entity .....</b>	<b>142</b>
7.2.1	Identita exempláře knihy jako zdroje digitalizátu .....	143
7.2.2	Identita vydání knihy jako zdroje digitalizátu .....	144
7.2.3	Identita digitalizátu knih jako intelektuální entity .....	145
<b>7.3</b>	<b>Identita a integrita objektu CDO .....</b>	<b>145</b>
<b>7.4</b>	<b>Řetězec odvozování objektu CDO .....</b>	<b>147</b>
7.4.1	Tištěná předloha jako první objekt CDO .....	147
7.4.2	Vytváření digitálního originálu .....	148
7.4.3	Dodávání digitálního originálu do archivu .....	151
7.4.4	Odvozování při archivaci .....	151
7.4.5	Odvozování při vytváření balíčku DIP .....	152
<b>7.5</b>	<b>Problémy procesů transformace objektu CDO .....</b>	<b>153</b>
7.5.1	Stanovení cílů transformace .....	153
7.5.2	Volba transformačních nástrojů .....	154
7.5.3	Validace výstupů transformace .....	155
7.5.4	Možnosti ověření věrnosti digitálního originálu k tištěné předloze .....	157
7.5.5	Znalost limitů a zkreslení transformací .....	158
<b>7.6</b>	<b>Otázky zpřístupňování .....</b>	<b>159</b>
7.6.1	Organizační řešení zpřístupňování .....	159
7.6.2	Prezentace jako způsob řešení problémů reprodukce .....	160
7.6.3	Možnost stažení jako doplňková forma zpřístupnění .....	161
7.6.4	Popisné informace prezentované spolu s digitalizátem .....	162
7.6.5	Autorizace digitální knihovny jakožto původce .....	163
<b>7.7</b>	<b>Opravy digitalizátů knih .....</b>	<b>163</b>
7.7.1	Opravy balíčku SIP před jeho dodáváním do archivu .....	164
7.7.2	Opravy při příjmu balíčku SIP do archivu .....	164
7.7.3	Opravy balíčků AIP v archivu .....	165
7.7.4	Opravy balíčku DIP .....	165
<b>7.8</b>	<b>Dokumentace a metadata .....</b>	<b>166</b>
7.8.1	Záznam atributů identity intelektuální entity .....	166
7.8.2	Záznam atributů integrity intelektuální entity .....	167
7.8.3	Záznam atributů identity objektu CDO .....	167
7.8.4	Záznam atributů integrity objektu CDO .....	168
7.8.5	Determinace metadatovými standardy .....	168
7.8.6	Aplikační metadatový profil a způsob plnění metadatových elementů .....	169
7.8.7	Kvalita bibliografických záznamů .....	170

7.8.8	Dokumentace standardů a nástrojů .....	171
<b>7.9</b>	<b>Shrnutí .....</b>	<b>172</b>
7.9.1	Proceduralita .....	172
7.9.2	Identita a identifikace .....	172
7.9.3	Validace jako kontrola integrity .....	173
7.9.4	Dostupnost a důvěryhodnost specializovaných nástrojů .....	174
7.9.5	Úplnost a rozsah dokumentace .....	174
7.9.6	Opravy jako aktuální problém .....	175
7.9.7	Komplexita obrazových dat .....	175
<b>8</b>	<b>NÁVRH POSTUPU PRO UDRŽENÍ AUTENTICITY DIGITALIZÁTŮ KNIH.....</b>	<b>176</b>
8.1	Obecný procedurální postup (sada „P-0“) .....	176
8.2	Obecný postup pro transformace objektu CDO (sada „P-1“) .....	178
8.3	Požadavky pro digitalizaci (sada „P-2“).....	181
8.4	Požadavky pro archivaci (sada „P-3“).....	187
8.5	Požadavky pro digitální knihovnu (sada „P-4“).....	190
<b>9</b>	<b>NÁVRH PRAVIDEL PERZISTENTNÍ IDENTIFIKACE DIGITALIZÁTŮ KNIH .....</b>	<b>192</b>
9.1	Základní požadavky pro perzistentní identifikaci digitálních objektů .....	192
9.1.1	Úvod.....	192
9.1.2	Obecné požadavky na perzistentní digitální identifikátory .....	194
9.1.3	Obecné požadavky na identifikační systémy .....	198
9.2	Požadavky na perzistentní identifikaci digitalizátů knih .....	201
9.2.1	Požadavky pro perzistentní identifikaci vydání tištěných knih .....	202
9.2.2	Požadavky pro perzistentní identifikaci digitalizátů knih.....	204
<b>10</b>	<b>ZÁVĚR.....</b>	<b>208</b>
	<b>SEZNAM POUŽITÝCH PRAMENŮ A LITERATURY .....</b>	<b>215</b>
	<b>SEZNAM ZKRATEK.....</b>	<b>225</b>
	<b>SEZNAM TABULEK V TEXTU .....</b>	<b>227</b>



# Předmluva

Tématem předkládané disertační práce je otázka autenticity digitalizovaných knih v celém jejich životním cyklu. Práce se zabývá požadavky na udržení autenticity digitalizátů knih v těch organizacích, které je jak vytvářejí (digitalizují), tak i následně uchovávají a zpřístupňují čtenářům jako součást digitálního dědictví.

V době, kdy jsem stanovil toto téma, jsem dokončoval magisterskou práci, která se věnovala základním principům digitální archivace (dlouhodobého uchovávání digitálních dokumentů).<sup>1</sup> Tato magisterská práce byla po menších úpravách později vydána knižně (1). V době jejího dokončování jsem zaznamenal, že jako jeden z dílčích problémů digitální archivace se intenzivně řeší otázka, jakými novými metodami a mechanismy zajistit udržení autenticity digitálních dokumentů. Některé zdroje dokonce považují otázku autenticity za ústřední problém digitální archivace. Uvedené skutečnosti byly základem mé motivace pro zvolení tématu disertační práce.

K pozdějšímu zúžení tématu na konkrétní typ dokumentu (digitalizát knihy) mě vedly dvě skutečnosti. Za prvé studie, v nichž odborníci poukazovali na to, že různé typy digitálních dokumentů vyžadují odlišné požadavky na autenticitu. Druhým faktorem byla skutečnost, že jsem se jako zaměstnanec specializovaného oddělení Národní knihovny ČR<sup>2</sup> pracovně zabýval otázkami dlouhodobého uchovávání převážně digitalizovaných knih. Toto pracovní prostředí pro mě představovalo nenahraditelný zdroj empirické zkušenosti, ze kterého jsem v této práci také do určité míry čerpal. Znalost některých specifických problémů životního cyklu, které popisují v této práci, jsem nezískal z žádné odborné literatury, ale jen díky nutnosti řešit konkrétní problémy archivační praxe, které byly často zcela neočekávané nebo překvapivé. Vzhledem k důležitosti přikládané tématu autenticity digitálních informací v odborné literatuře a ke skutečnosti, že dosud neexistuje soustavná konceptualizace požadavků autenticity digitalizátů knih, lze konstatovat, že zvolené téma disertační práce je vysoce aktuální.

Cíle disertační práce jsou konceptualizace požadavků autenticity v kontextu digitalizátů knih a doporučený postup pro udržení autenticity digitalizátů knih pro celý jejich

---

<sup>1</sup> Tuto magisterskou práci jsem na podzim roku 2009 na Ústavu informačních studií a knihovnictví Filozofické fakulty Univerzity Karlovy v Praze úspěšně obhájil.

<sup>2</sup> Od roku 2008 pracuji v Oddělení pro standardy Národní knihovny ČR, mezi jehož úkoly patří vytváření národních standardů pro digitální archivaci.

životní cyklus. Součástí doporučeného postupu je návrh pravidel pro perzistentní identifikaci tohoto typu dokumentu.

Pro naplnění výše uvedených cílů bylo třeba vytvořit přehled konceptů normy ISO 14721 (jako základního konceptuálního rámce pro řízení životního cyklu jakýchkoli digitálních dokumentů), zpracovat přehled současného životního cyklu digitalizátů knih v praxi knihoven (zahrnujícího jak digitalizaci, tak dlouhodobé uchovávání a zpřístupňování) a provést analýzu a revizi relevantních tematizací autenticity digitálních dokumentů.

Na základě výstupů z výše uvedených kroků jsem vytvořil obecný rámec pro zkoumání autenticity, který jsem aplikoval na sledování specifických souvislostí mezi autenticitou a řízením životního cyklu digitalizátů knih. Výsledkem této aplikace je konceptualizace požadavků na udržování a posuzování autenticity digitalizátů knih, která obsahuje identifikaci největších problémů této oblasti a která představuje plnění prvního cíle této disertační práce. Na základě této konceptualizace jsem pak vytvořil doporučený postup pro udržení autenticity v životním cyklu digitalizátů knih, který se zaměřuje na řešení některých problémů autenticity. Tento postup je dále rozpracován do jedné z dílčích oblastí řešení požadavků autenticity (oblasti perzistentní identifikace), a to v podobě návrhu pravidel pro perzistentní identifikaci digitalizátů knih. Tento návrh byl částečně uplatněn v praxi. Některé jeho body jsem rozpracoval do pravidel systému ČIDLO (Český IDentifikační a LOkalizační systém) a dále do návrhu některých funkcí jeho softwarového řešení, jehož vývoje jsem byl garantem (2).

Výsledky práce jsou prezentovány v devíti kapitolách. Kapitola 2 shrnuje základní konceptuální rámec pro řízení životního cyklu digitálních dokumentů z perspektivy dlouhodobého uchovávání, který je široce uznáván a ze kterého vycházím v této práci i já. Tato kapitola prezentuje a vysvětluje základní pojmy a modely normy ISO 14721. Uvádí také některé souvislosti se specifickým kontextem knihoven. Vzhledem k tomu, že uvedená norma je všeobecně přijímána, převzetí její základní terminologie v této práci není třeba zdůvodňovat.

Kapitoly 3 a 4 představují systematický přehled řízení životního cyklu digitalizátů knih v současných knihovnách. Kapitola 3 pojímá digitalizát knihy ze statického hlediska (jako popis jeho složek, doprovodných metadat a souvisejících standardů a nástrojů). Kapitola 4 zohledňuje dynamické hledisko, tedy hlavní etapy životního cyklu digitalizované knihy (produkce, archivace, zpřístupňování). Obě kapitoly slouží jako shrnující popis předmětu autenticity (tj. digitalizátů knih) a řízení jejich životního cyklu v současné praxi, s cílem

zobecnit klíčové aspekty a principy praxe. Jako zdroje jsou užity zejména směrnice, standardy a specializované nástroje a dále případy z konkrétních projektů.

Kapitola 5 podává přehled současných tematizací autenticity digitálních dokumentů, které jsou relevantní pro digitalizáty knihy, a přináší jejich předběžnou analýzu. Základem kapitoly je popis konceptualizace autenticity elektronických archiválií, která se jako jediná ze zdrojů specificky zabývá autenticitou určitého typu dokumentu. Tato archivářská konceptualizace se stala příkladem pro plnění druhého cíle disertační práce. Kapitola dále popisuje vybrané další tematizace autenticity z oblasti digitálního dědictví. Žádný z těchto zdrojů nepojednává výhradně o autenticitě, nýbrž autenticitu jen zmiňuje v souvislosti s různými aspekty životního cyklu.

Kapitola 6 je revizí uvedených tematizací. Jejím cílem je identifikovat společné prvky a nastínit možnou škálu okruhů zkoumání. Na základě výstupů revize jsem pak na konci této kapitoly formuloval obecný rámec pro zkoumání autenticity digitalizátů knih.

V Kapitole 7 jsem využil tohoto obecného rámce ke specifikaci požadavků autenticity při produkci, archivaci a zpřístupňování digitalizovaných dokumentů. Výstupy této kapitoly tedy plní první cíl této disertační práce.

Kapitola 8 na základě konceptualizace uvedené v předchozí kapitole navrhuje doporučený postup pro udržení autenticity v životním cyklu digitalizátů knih. Tento postup se skládá ze dvou částí: první část navrhuje obecná pravidla pro celý cyklus, druhá část navrhuje postupy pro jednotlivé fáze. Navržený postup pochopitelně neřeší všechny problémy popsané v konceptualizaci Kapitoly 7. Takový cíl přesahuje možnosti individuální výzkumné činnosti.

Kapitola 9 rozšiřuje doporučený postup o návrh pravidel perzistentní identifikace digitalizátů knih. Perzistentní identifikaci lze chápat jako jedno z řešení problému, jak udržet autenticitu v celém životním cyklu. Tento návrh se snaží zohlednit všechny hlavní aspekty perzistentní identifikace ve vztahu k autenticitě. Část těchto pravidel se pak stala základem pro můj návrh systému ČIDLO, který jsem vytvořil pro Národní knihovnu ČR v letech 2011-2013 (2) a který je dále vyvíjen. Úspěšné fungování systému ČIDLO je možné považovat za určitý doklad praktické aplikovatelnosti návrhu obsaženého v Kapitole 9. Systém ČIDLO přidělil digitalizovaným knihám (a dalším typům dokumentu) více než jeden milion identifikátorů URN:NBN, což z něj patrně činí největší identifikační systém dokumentů digitálního dědictví v ČR. Z analýzy evropských národních identifikačních

systémů URN:NBN také vyplývá, že český systém patří z hlediska pravidel a funkcí mezi nejkomplexnější (3).

Formální úprava disertační práce vychází z pokynů Filozofické fakulty Univerzity Karlovy. Použité prameny a literatura jsou citovány v souladu s normami ČSN ISO 690 a ČSN ISO 690-2, pro citace v textu je použita číselná metoda (v případech, kdy je citována konkrétní část daného dokumentu, např. konkrétní stránka, následují tyto údaje po číslu odkazujícím na položku seznamu použitých pramenů a literatury).

Některé výsledky uvedené v této práci již byly částečně publikovány. Části podkapitoly 9.1 byly původně publikovány v recenzovaném časopise Knihovna (4) a jako technická zpráva pro projekt Správa elektronických publikací v síti knihoven ČR (5). Pro tuto disertaci jsem své příspěvky přepracoval a rozšířil. V letech 2012-2013 jsem byl také řešitelem vývoje softwaru<sup>3</sup> pro systém ČIDLO (2), a jako garant projektu i autor konceptuálního návrhu; hlavním vývojářem byl Martin Řehánek. Konečná podoba systému ČIDLO byla stručně prezentována v přehledovém článku (6). V roce 2014 jsem se jako spoluautor podílel na zpracování pravidel systému ČIDLO do formální podoby, která byla 5. června 2015 uznána Ministerstvem kultury ČR jako certifikovaná metodika (7).

Za poskytnutí cenných rad a konzultací v průběhu zpracovávání disertační práce a také za trpělivost bych v první řadě rád poděkoval svému školiteli prof. RNDr. Jiřímu Ivánkovi, CSc. Dále bych rád poděkoval také PhDr. Lindě Jansové, Ph.D. a všem kolegům, kteří mi poskytli zajímavé podněty nebo připomínky z praxe.

---

<sup>3</sup> V rámci Koncepce rozvoje Národní knihovny České republiky jako výzkumné organizace od r. 2010 do r. 2015, Oblast 2, Priorita 2.3: Vývoj podpůrných / návazných aplikací pro dlouhodobou ochranu digitálních dat



# 1 Úvod

Ke skutečnosti, že autenticita je jedním z klíčových problémů digitální archivace a že téma autenticity v digitálním světě jsem dospěl již během studia materiálů pro zpracování mé magisterské práce. Jedním z rozhodujících zdrojů pro mě byla úvodní studie shrnující problematiku digitální archivace, kterou v roce 2008 publikovala Priscilla Caplanová, jedna z hlavních odbornic v oboru a spoluautorka metadatového standardu PREMIS (8). Caplanová ve studii představuje sedm cílů digitální archivace, z nichž nejvyšším cílem je právě udržení autenticity digitálních dokumentů. V té době byl již také zveřejněn návrh druhého vydání normy ISO 14721, jehož hlavní změny zahrnovaly právě nově definovaný pojem autenticity.<sup>4</sup>

K zúžení tématu na autenticitu určitého typu digitálního dokumentu mě vedly studie, které argumentovaly ve prospěch konkretizace výzkumu autenticity na základě specifik různých typů objektů a kontextů užití.<sup>5</sup> Oblast klasické hudby například tematizuje otázku autenticity v souvislosti se vztahem mezi původní hudební partiturou a jejími interpretacemi různými hudebními tělesy v průběhu let. Oblast digitalizovaných filmů v českém prostředí aktuálně řeší otázku, jak restaurovat filmový materiál do původní (autentické) podoby, která vyvolává spory o autenticitu mezi Národním filmovým archivem a Asociací českých kameramanů.

V této práci je předmětem zkoumání digitalizovaná kniha v kontextu její správy jednou organizací po celý životní cyklus, kdy všechny etapy má pod kontrolou jedna organizace. Vyšší míra kontroly ale také předpokládá vyšší míru odpovědnosti, a tedy i vyšší míru požadavků na autenticitu. Knihovny již tím, jakým způsobem vytvářejí digitalizáty knih, do určité míry ovlivňují možnosti jejich archivace a zpřístupňování. Tento problém je poměrně široce známý a zejména (ale nikoliv výlučně) národní knihovny zpravidla kladou velký důraz na takovou produkci digitalizovaných knih, která usnadňuje vyhlídky jejich následné digitální archivace (jde například o volbu vhodného formátu, u kterého nehrozí jeho brzké zastarání). Základním kontextem autenticity pro tuto disertační práci je tedy úplné a komplexní řízení životního cyklu digitalizátu knih v knihovně, která je současně producentem digitalizované knihy i archivujícím a zpřístupňujícím subjektem. Takový kontext existuje již od počátku masové digitalizace a nejedná se tedy o výjimečný ideální případ.

---

<sup>4</sup> Srv. (1, s. 131).

<sup>5</sup> Viz zejména kniha H. Gladneyho (90, s. 93-96) a studie D. Bearmana a J. Trantové (86).

Pro naplnění cíle disertační práce jsem od začátku zvolil postup analýzy směřující od obecného ke konkrétnímu. Tento postup zahrnoval tyto navazující zužující kroky: a) zvolení základního konceptuálního rámce pro popis životního cyklu digitálních dokumentů obecně, b) systematický popis současné praxe životního cyklu digitalizátů knih jako specifického kontextu, který je výchozím prostředím pro zkoumání otázky autenticity v této práci, c) analýza tematizací autenticity, které jsou relevantní pro plnění cílů disertační práce, d) revize těchto pojetí a vytvoření obecného rámce pro zkoumání autenticity užitím výstupů této revize, e) vytvoření konceptualizace požadavků autenticity digitalizovaných knih (uplatněním tohoto obecnějšího rámce na okruhy současné praxe), e) navržení doporučených postupů vycházejících z této konceptualizace.

Jako základní konceptuální rámec pro výzkum byl zvolen model OAIS obsažený v normě ISO 14721 (10). Tento model je již od počátku milénia základním a obecně přijímaným referenčním rámcem pro řízení životního cyklu digitálních dokumentů z perspektivy potřeb digitální archivace. Z tohoto důvodů je jeho užití zcela opodstatněné. Tím je stanoven univerzální výchozí bod (jednotná a srozumitelná terminologie a definice základních požadavků a postupů digitální archivace) pro další specifikaci požadavků autenticity digitalizátů knih. Model OAIS se primárně zabývá archivací a zpřístupněním, ale do určité míry vymezuje i otázku produkce digitálních dokumentů. Činí tak v podobě konceptu dohody o dodávání dat, která je vyjádřením toho, v jaké podobě budou digitální objekty dodávány do archivu, a výsledkem dohody mezi producentem a archivem. V praxi může docházet k tomu, že archiv nemá možnost ovlivnit tuto podobu způsobem, který odpovídá požadavkům digitální archivace, například pokud má legislativou stanovené povinnosti přijímat a archivovat digitální publikace od vydavatelů a současně platí, že legislativa vydavatelům neukládá povinnost dodávat digitální publikace do archivu ve formátech, které archiv preferuje, a archiv nemá možnost provádět formátovou konverzi publikací do požadované podoby (např. z důvodu neexistence adekvátních nástrojů pro takovou konverzi). V této práci se však zabýváme optimálním případem, kdy taková možnost existuje, jelikož je producentem, vkladatelem i archivem jedna a tatáž organizace (a tudíž může jako cílové formáty digitalizace zvolit formáty vhodné pro archivaci), a kdy vhodné technologie jsou dostupné (tj. zejména existují vhodné formáty).

Během přípravné fáze, kterou jsem věnoval studiu odborné literatury a dalších relevantních zdrojů (např. digitalizačních směrnic), jsem našel pouze jeden zdroj, který přináší soustavnou konceptualizaci specifického typu digitálního dokumentu a který pochází z příbuzné oblasti. Jedná se o výstupy projektu InterPARES, který byl přímo zaměřen

na otázky autenticity elektronických archiválií. Výstupy projektu jsou systematické a koherentní, ale je třeba podotknout, že projekt se nezabýval specifickými otázkami různých dílčích typů elektronických archiválií, ale hledal obecným model platným pro všechny archiválie. Žádný z dalších zkoumaných zdrojů z oblasti knihovnictví nebo příbuzných oblastí se autenticitou specificky nezabývá – ve všech případech je otázka autenticity pouze jedním z několika témat, o kterých zdroj pojednává. V roce 2012 byla vydána revidovaná verze normy ISO 14721 (ISO 14721:2012), která přinesla velmi obecný rámec pro autenticitu digitálních dokumentů – v tom smyslu, že uvedla definici pojmu autenticita a dvou souvisejících konceptů.

Výše uvedené skutečnosti mě vedly k postupu, kdy jsem jako základní zdroj pro otázky konceptualizace autenticity digitalizátů knih užil obecný rámec pro autenticitu daný normou ISO 14721:2012, s tím, že jej nebudu nijak měnit a pouze jej dále rozšířím. Základem pro toto rozšíření se stal model projektu InterPARES, který jsem převzal a upravil. Dále jsem využil teze o různých dílčích aspektech autenticity obsažené v dalších zdrojích, o kterých uvedené hlavní zdroje (tj. norma ISO 14721:2012 a projekt InterPARES) nepojednávaly (nebo o nich pojednávaly pouze obecně) a které jsem shledal za relevantní pro téma disertační práce.

Součástí celého procesu byla kritická a komparativní analýza vybraných tematizací a definic. Nebylo možno převzít existující modely a definice a užít je jako základ pro specifikaci konceptualizace autenticity digitalizátů knih. Pro účely analýzy tak bylo klíčové vytvoření obecného rámce pro zkoumání autenticity. Tento obecný rámec a závěry z revize tematizací tak společně představují dílčí výsledek výzkumu, který je dále využit pro plnění hlavních cílů práce. Má však i určitý význam sám o sobě (např. jako naznačení širší problematiky).

Při analýze problému jsem také bohatě čerpal z empirických poznatků vlastní praxe, které se staly neocenitelným zdrojem informací, a to ve dvou oblastech. Zaprvé v oblasti případů, které jsou tak specifické, že o nich zkoumané odborné zdroje nepojednávají. Zadruhé v oblasti identifikace chyb a problémů, které se v konceptualizacích buď z principu nedají předvídat, nebo o nich zdroje neinformují.<sup>6</sup> Okrajově jsem užil i vlastního testování některých nástrojů s cílem prakticky ověřit jejich fungování na vlastním datovém vzorku a zjistit tak

---

<sup>6</sup> Důvodem může být i skutečnost, že paměťové instituce, které často bojují o samotné uznání potřeby zvýšit financování vzhledem k novým technologickým nárokům, nemají ve svém zájmu prezentovat nejrůznější chyby, v obavě, že by tyto chyby mohly být vykládány jako jejich neschopnost, ačkoliv se jedná o skutečně nepředvídatelné problémy.

možnosti, resp. omezení, která tyto nástroje kladou, a tím nutně vytvářejí determinující technologické prostředí, které limituje výstupy současných postupů. Všechny tyto zdroje empirické zkušenosti jsem se pokusil zobecnit, aby měly nadčasový rámec a nepopisoval jsem detaily, které v blízké budoucnosti pravděpodobně přestanou být aktuální (např. chyby nástrojů, které v další verzi budou opraveny).

Konceptualizace požadavků autenticity digitalizátů knih se ukázala být velmi komplexní. Doporučený postup pro udržení autenticity digitalizátů knih, který z ní vychází, nepokrývá všechny problémové okruhy uvedené v této v konceptualizaci – takový úkol se ukázal být nad rámec rozsahu této práce i nad možnosti individuálního výzkumu. Mnou uvedená konceptualizace však může posloužit například jako zdroj informací, kudy by se mohly ubírat další, specializovanější výzkumy (například v oblasti převodů specifických obrazových vlastností, jako je barevný profil). Návrh pravidel perzistentní identifikace digitalizátů knih si klade za cíl pokrýt všechny základní aspekty tohoto způsobu řešení, jak udržet a posoudit autenticitu digitalizátů knih v dlouhodobém horizontu a v internetovém prostředí. Za jistý doklad jeho životaschopnosti lze považovat jeho implementaci do návrhu pravidel pro systém perzistentní identifikace ČIDLO, který jsem vytvořil pro Národní knihovnu ČR v letech 2011-2013 a který úspěšně funguje (6). Tento návrh byl později rozšířen (např. o popis postupu konkrétní implementace, jako je způsobu plnění metadat) a zpracován do podoby certifikované metodiky, jejímž jsem spoluautorem (7).

## **2 Základní konceptuální rámec životního cyklu**

Základním konceptuálním rámcem pro životní cyklus digitálních dokumentů (tedy nejen digitalizátů knih, ale i všech dalších typů těchto dokumentů) v organizacích, které usilují o dlouhodobé uchovávání a zpřístupňování digitálního dědictví, je jednoznačně norma ISO 14721. Hlavním obsahem této normy je model pro digitální archivaci (model OAIS). Tento model zahrnuje nejen otázku uchovávání, ale také zpřístupňování a do určité míry i produkce digitálních dokumentů. První vydání normy ISO 14721 vyšlo v roce 2003 (9), druhé v roce 2012 (10). Od přelomu milénia (kdy již byl znám návrh modelu OAIS) se prakticky všechny významné knihovny světa i celá řada jiných organizací rozhodly z modelu OAIS vycházet a na jeho základě navrhovat nebo budovat svá konkrétní archivační řešení. Druhou nejdůležitější normou v oblasti digitální archivace je norma ISO 16363 (11). Tato norma je všeobecně přijímána jako nástroj pro zjišťování míry souladu archivačního systému konkrétní organizace s modelem OAIS obsaženým v normě ISO 14721, a to za účelem vnitřního auditu i oficiální certifikace konkrétního archivu externím subjektem.

### **2.1 Historie norem**

Odborná komunita v 90. letech dospěla k poznání, že nárůst užívání digitálních technologií, jejich rychlé změny a zvyšující se komplexita digitálních objektů znamenají nový typ problému pro dlouhodobé uchovávání digitálních objektů. Bylo shledáno, že je třeba vytvořit konceptuální rámec, který by systematicky uchopil tuto novou problematiku. Mezinárodní organizace pro normalizaci (International Organization for Standardization, ISO) požádala Poradní výbor pro systémy pro data z kosmického prostoru (Consultative Committee for Space Data Systems, CCSDS), organizaci sdružující kosmické agentury, o vytvoření normy pro digitální archivaci. V lednu 2002 publikoval výbor CCSDS konečnou verzi dokumentu, a to jako svůj standard CCSDS 650.0-B-1, známý jako standard OAIS (Open archival information system) (12). Organizace ISO tento standard převzala a po formální úpravě vydala jako normu ISO 14721:2002 s názvem „Space data and information transfer systems – Open archival information system (OAIS) – Reference model“ (Systémy pro přenos dat a informací z kosmického prostoru – Otevřený archivační informační systém – Referenční model) (9).

Norma prošla v průběhu let klasickým revizním řízením, které vedl opět výbor CCSDS a jeho výsledek vydal jako novou verzi původního standardu pod označením CCSDS

650.0-M-2 (13). Tento revidovaný standard převzala a po formálních úpravách vydala organizace ISO v roce 2012 jako druhé vydání normy ISO 14721 (10) .

Je třeba odlišovat standard výboru CCSDS a normu ISO 14721 vydanou organizací ISO. Standard CCSDS je volně dostupný online na webovém sídle výboru CCSDS, zatímco normu ISO 14721 je nutno zakoupit. Norma ISO má také jiný rozsah (obsahuje navíc formální části vztahující se k organizaci ISO) a zahrnuje číselnou paginaci. Vlastní text obou dokumentů je však totožný. Ve většině zahraničních prací je citován standard CCSDS, nikoliv příslušná norma ISO.<sup>7</sup>

Norma ISO 14721:2003 obsahovala také plán pro přípravu souvisejících norem. Součástí tohoto plánu byla norma pro certifikaci archivů OAIS. V návaznosti na tuto skutečnost vypracovala pracovní skupina organizací Sdružení vědeckých knihoven (Research Libraries Group; RLG) a NARA dokument, který byl vydán v roce 2007 a je znám pod zkratkou „TRAC“ (Trustworthy Repositories Audit & Certification: Criteria and Checklist; Audit a certifikace důvěryhodných úložišť: kritéria a kontrolní seznam) (14). Knihovny začaly dokument TRAC užívat zejména pro vnitřní audit svých archivů. Tento dokument převzal výbor CCSDS, aby jej přepracoval do návrhu normy ISO a v roce 2011 vydal jako svůj standard pod označením CCSDS 652.0-M-1 (15). Standard poté po obvyklých formálních úpravách vydala organizace ISO jako normu ISO 16363:2012 pod názvem „Audit and certification of trustworthy digital repositories“ (Audit a certifikace důvěryhodných digitálních úložišť)“ (11). Je třeba upozornit, že dokument TRAC se od obsahu standardu CCSDS 652.0-M-1 i normy ISO 16363:2012 poměrně výrazně liší. Část pasáží byla sice zachována, ale větší část byla přeformulována nebo nahrazena novými.

Český Úřad pro technickou normalizaci, metrologii a státní zkušebnictví (ÚNMZ) vydal české verze mezinárodních norem ISO 14721:2012 (16) a 16363:2012 (17). Na překladu těchto norem jsem se podílel, nicméně v této disertační práci cituji mezinárodní normy ISO 14721:2012 a ISO 16363:2012 (tj. anglické verze) v překladu, který se většinou od českého znění norem liší. Důvodem je zachování celkové konzistence české terminologie užití v této disertační práci a odlišné překladatelské požadavky.

---

<sup>7</sup> V mé knize (1) byla také citována první verze standardu CCSDS, nikoliv norma ISO vydaná v roce 2003.

## 2.2 Norma ISO 14721

Norma ISO 14721 ve své nejobecnější rovině popisuje dlouhodobé uchovávání jakéhokoliv typů informací (digitálních i nedigitálních), blíže však popisuje pouze digitální archivaci (dlouhodobé uchovávání digitálních informací). Tato norma uvádí, že „poskytuje rámec, který může být dále rozšířen, aby pokrýval oblast dlouhodobého uchovávání informací, jež se nevyskytují v digitální formě (např. fyzických nosičů nebo fyzických vzorků)“ (10, s. 1).

### 2.2.1 Úvodní přehled

Předmětem normy ISO 14721 je model pro otevřený archivační informační systém (model OAIS), který se skládá ze dvou dílčích modelů, jimiž jsou funkční model (*functional model*) a informační model (*information model*). Široce známým a zmiňovaným v odborné literatuře i v knihovnické praxi je spíše funkční model. Funkční model popisuje funkce otevřeného archivačního informačního systému (archivu OAIS). Těžištěm zájmu této dizertační práce je však informační model, který přináší klasifikaci různých typů informačních objektů.

Příčinou přetrvávající redukce modelu OAIS na funkční model je zřejmě skutečnost, že na základě funkčního modelu lze navrhovat, budovat a vyvíjet softwarový systém podporující potřeby digitální archivace. Z tohoto důvodu stále dochází k tomu, že knihovny věnují informačnímu modelu OAIS menší pozornost, než by si zasloužoval. Na toto opomíjení upozorňuje i související norma ISO 16363:2012 (11, s. 15). Upozadění informačního modelu v současné praxi přináší zásadní implementační problém, protože konceptuálně je nejprve nutné přesně definovat objekt uchovávání, a teprve poté lze zavést systém určený na uchovávání tohoto objektu.

Důraz na funkční model OAIS je také jednou z možných příčin skutečnosti, že se v praxi pojem archiv OAIS někdy mylně ztotožňuje se softwarovým systémem a vyskytuje se mylná představa, že softwarový systém sám o sobě může naplňovat všechny požadavky modelu OAIS. Archiv OAIS však nezahrnuje pouze softwarový systém, ale také kompletní technologické prostředí (hardware, datové nosiče, síťové technologie, záložní zdroje atd.), fyzické prostory, lidské zdroje nebo organizační zajištění. Toto komplexní pojetí archivu OAIS, které norma přináší, přitom odpovídá klasickému pojetí informační vědy. Připomeňme známou definici Briana Vickeryho ze 70. let 20. století, podle níž je informační

system „organizace lidí, materiálů a strojů, která slouží k usnadnění přenosu informací od jedné osoby k druhé“ (18, s. 1).

### 2.2.1.1 Pojetí informací

Pojem „informace“ (*information*) definuje norma ISO 14721 jako „jakékoliv znalosti (*knowledge*), které mohou být předmětem výměny (*exchange*)“ a udává, že při výměně jsou informace „vždy vyjádřeny (tj. reprezentovány) určitým typem dat“ (10, s. 30). Data jsou definována jako „opakovaně interpretovatelná reprezentace informací ve formalizované podobě vhodné pro komunikaci, interpretaci nebo zpracování.“ (10, s. 21). Pojmy informace a data jsou kategorie. Jednotlivý objekt spadající do první kategorie nazývá norma ISO 14721 informační objekt (*information object*), objekt druhé kategorie datový objekt (*data object*).

Při výměně tedy příjemce získává informace vždy z dat, v tom smyslu, že převádí datový objekt na informační objekt. Aby se tento proces mohl uskutečnit, musí příjemce disponovat odpovídající znalostní základnou (*knowledge base*), což je „množina informací, které si osvojila osoba nebo systém a která této osobě nebo tomuto systému umožňuje porozumět přijímaným informacím“ (10, s. 23).

Porozumění informací příjemcem je tedy chápáno jako převod datového objektu na informační objekt užitím znalostní základny příjemce. Pokud si například český čtenář zapůjčí knihu „Information Science in Theory and Practice“ Briana Vickeryho, pak pozorovatelné znaky (vytištěná slova) představují datový objekt. Aby příjemce knize rozuměl, tj. mohl převést tyto znaky na informace, musí umět anglicky. Pokud tomu tak je, pak to znamená, že znalost angličtiny tvoří součást jeho znalostní základny. Čtenář, který anglicky neumí, bude potřebovat získat dodatečné informace, aby knize rozuměl. Tento typ informací norma OAIS označuje za interpretační informace (*representation information*) a definuje je jako „informace, které převádějí datový objekt do smysluplnějších významových celků (*the information that maps a data object into more meaningful concepts*)“ (10, s. 25). Tato definice působí poněkud krkolomně, nicméně jejím smyslem je patrně poukázat na skutečnost, že srozumitelnost lze posuzovat ve stupních. Ve výše uvedeném příkladu i čtenář, který anglicky neumí, rozumí tomu, že text je v angličtině; pokud by nevěděl, že text je v angličtině, mohl by rozumět alespoň tomu, že text je v cizím jazyce; na nejnižším stupni (např. pokud je negramotný) pak může stále rozumět alespoň tomu, že datový objekt je kniha. Interpretační informace jsou tedy informace, které popisují formu reprezentace informací v datech (v našem příkladu je jí anglický jazyk). Tyto informace potřebuje získat ten, jehož znalostní základna informace o formě reprezentace neobsahuje. Čtenář, který neumí anglicky,



si například může opatřit českou učebnici angličtiny a anglicko-český slovník, což jsou datové objekty, ze kterých již na základě své znalostní základny dokáže získat informace: učebnice a slovník tedy čtenáři poskytují interpretační informace, které potřebuje pro porozumění anglicky psané knize. Můj český překlad „interpretační informace“ není doslovný (ten by byl „informace o reprezentaci“). Podle mého soudu však lépe vystihuje skutečnost, že veškerá data musejí být předmětem interpretace.

Datový objekt norma ISO 14721 definuje výčtem – datový objekt je „buď fyzický objekt, nebo digitální objekt“ (10, s. 21). Fyzický objekt (*physical object*) je „objekt (například měsíční hornina, biologický vzorek, mikroskopické sklíčko) s fyzicky pozorovatelnými vlastnostmi, které reprezentují informace, jež je pro účely uchovávání, šíření a samostatného využívání vhodné patřičně zdokumentovat“ (10, s. 24). Digitální objekt (*digital object*) je „objekt složený z množiny bitových posloupností (*a set of bit sequences*)“ (10, s. 22). Informační objekt (*information object*) je pak „datový objekt se svými interpretačními informacemi“ (10, s. 23).

Norma ISO 14721 nedefinuje slovní spojení „digitální informace“ (*digital information*), ačkoliv jej často užívá, ale z textu lze vyvodit, že digitální informace jsou informační objekty, které vznikly interpretací digitálních objektů. Pro označení informací vzniklých interpretací fyzických objektů užívá norma termín „nedigitální“ (*non-digital*).

Již tyto úvodní definice ukazují, že pojetí informací v normě ISO 14721 se liší od řady jiných modelů. Například podle modelu projektu InterPARES 2 jsou informace (*information*) „sdělení zamýšlené pro komunikaci napříč časem a prostorem“, data (*data*) jsou „nejmenší smysluplná část informace“ a dokument (*document*) je „zaznamenaná informace“ (19, s. 15). Data jsou tedy podle modelu podkategorií informací (zatímco v normě ISO 14721 jsou „data“ a „informace“ dvě odlišné kategorie stejné úrovně) a definice termínu „dokument“ odpovídá „datovému objektu“ v normě OAIS.

### 2.2.1.2 Archiv OAIS

Norma ISO 14721 pojednává o specifickém modelu archivu, který označuje jako „otevřený archivační informační systém“ (*open archival information system*), zkráceně archiv OAIS. Definice archivu OAIS je pak následující: „archiv (*archive*), který se sestává z organizace tvořené lidmi a systémy, jež přijala odpovědnost uchovávat informace a zpřístupňovat je cílové komunitě (*designated community*), přičemž tento archiv může být součástí větší organizace.“ (10, s. 24). Norma ISO 14721 odlišuje „archiv OAIS“ od jiných užití pojmu „archiv“ pomocí sady definovaných povinností. Norma ISO 16363, která

se zabývá výlučně dlouhodobým uchováváním digitálních informací, zavádí pro termín „digitální archiv“ (ve smyslu archivu OAIS, který vykonává digitální archivaci) synonyma „repozitář (*repository*)“ nebo „digitální repozitář (*digital repository*)“ (11, s. 16). Norma ISO 16363 pak v textu nejčastěji užívá pojem „repozitář“. Výraz „otevřený“ vyjadřuje skutečnost, že doporučení normy a budoucí související doporučení vznikají ve veřejných diskusích; neznamená, že přístup k archivu je neomezený (10, s. 12).

Cílová komunita (*designated community*) je „stanovená skupina možných koncových uživatelů, kteří by měli být schopni porozumět konkrétní množině informací. Cílová komunita může být složena z více uživatelských komunit; cílovou komunitu si vymezuje archiv a její vymezení se v průběhu času může měnit.“ (10, s. 22). To znamená, že archiv OAIS musí při stanovení cílové komunity brát v potaz její aktuální znalostní základnu a na základě ní zvolit odpovídající množinu potřebných interpretačních informací, které bude muset uchovávat spolu s digitálními objekty. Vzhledem k tomu, že znalostní základna cílové komunity se může (a v praxi s nejvyšší pravděpodobností bude) v průběhu času měnit, mohou se postupně měnit i požadavky na množinu potřebných interpretačních informací. Pokud například bude archiv OAIS uchovávat textové dokumenty v jazyce, který je málo užíván, pak se může v okamžiku, kdy zjistí, že tento jazyk již začíná vymírat, rozhodnout shromáždit a uložit další interpretační informace, které umožní cílové komunitě rozumět tomuto jazyku i poté, co zanikne. Pokud by archiv OAIS k tomu opatření nepřistoupil včas, mohou se interpretační informace stát nedostupnými a z dokumentů v zaniklém jazyce již čtenáři nebudou moci získat dostatečně smysluplné informace.

Výše uvedená definice archivu OAIS však na první pohled působí zvláště, vzhledem k tomu, že norma ISO 14721 se zabývá dlouhodobým uchováváním, a přitom v této klíčové definici se hovoří pouze o uchovávání, nikoliv o dlouhodobém uchovávání. Vysvětlení, které nemusí být při četbě normy ISO 14721 na první pohled zřejmé, spočívá v tom, že podle této normy archiv OAIS udržuje „informace určené k dlouhodobému uchovávání, a to i v případě, kdy samotný archiv OAIS nebude fungovat trvale“ (10, s. 12). Archiv OAIS by proto podle této normy měl mít oficiální plán nástupnictví (*succession plan*) pro předání archivovaných dokumentů jinému archivu OAIS v případě, že přestane vykonávat svou činnost nebo řídicí nebo financující subjekt podstatně změni předmět své činnosti (10, s. 44). Je třeba také dodat, že plán nástupnictví je novým konceptem druhého vydání normy ISO 14721. Tento koncept se předtím vyskytoval ve standardu TRAC (2007) jako součást jednoho z požadavků důvěryhodnosti archivu (14, s. 10) a posléze byl převzat také do normy ISO 16363 (11, s. 22).

Norma ISO 14721 definuje dlouhodobé uchovávání (*long term preservation*) jako „dlouhodobé udržování informací – v podobě, která je určené skupině srozumitelná sama o sobě (*independently understandable*) –, a dokladů o jejich autenticitě“ (10, s. 24). Aspekt dokladů o autenticitě je další novinkou druhé verze, o které blíže pojednává oddíl 5.3.

Pojem „srozumitelný sám o sobě“ je definován jako „vlastnost informací, které jsou dostatečně úplné, aby jim určená skupina rozuměla a uměla je využít, aniž by bylo nutné se uchýlit ke specializovaným zdrojům, jež nejsou běžně dostupné, a to včetně konkrétních jednotlivců“ (10, s. 23). Pokud vyjdeme z předchozího příkladu s mizejícím jazykem, pak archiv OAIS bude interpretační informace potřebné k porozumění vymírajícímu jazyku shromažďovat a uchovávat pouze v případě, že tyto interpretační informace nebudou všeobecně běžně dostupné. Případ konkrétních jednotlivců zřejmě odkazuje například na výzkumníky, kteří vytvořili datové sady v rámci specifického výzkumu, jehož popis musí být spolu s datovými sadami uchováván také.

Koncepty nezávislé srozumitelnosti a znalostní základny (na straně cílové komunity) a interpretačních informací (na straně archivu) tak tvoří základ pro problém převodu uchovávaných digitálních objektů do informačních objektů.

Norma ISO 16363 upravuje termín „cílová komunita“ tak, že tento termín podle ní může zahrnovat více cílových komunit (11, s. 16): „Zatímco jeden repozitář může mít jednu zobecněnou „cílovou komunitu“ (např. každého občana dané země), jiné repozitáře mohou mít několik odlišných cílových komunit s úzce vymezenými potřebami a každá z nich může od repozitáře vyžadovat odlišné funkce nebo podporu.“ V kontextu knihoven by prvnímu případu odpovídala například národní knihovna („každý čtenář dané země“); druhému případu ale vlastně také může odpovídat národní knihovna, pokud nabízí odlišné funkce například pro badatele na straně jedné a dětské čtenáře na straně druhé. V každém případě ani toto rozšíření pojmu „cílová komunita“ zřejmě nezahrnuje případy, případy, kdy jeden archiv uchovává informace různých organizací, z nichž každá může teoreticky představovat jinou cílovou komunitu ve vztahu ke své podmnožině informací (reálným příkladem může být digitální archiv Portico, který uchovává články odborných periodik různých knihoven).

Cílová komunita archivu musí být podle normy ISO 14721 vždy nějak stanovena. A to i tehdy, pokud aktuálně neexistují žádní uživatelé informací uchovávaných v archivu. Příkladem takové praxe je tzv. „nepřístupný archiv (dark archive)“, jehož úkolem je dlouhodobě uchovávat informace, ale v krátkodobém horizontu je nikomu nepřístupňovat

(často z důvodu omezení vyplývajících z práv duševního vlastnictví). Uchovávání je v tomto případě tedy zaměřeno na možné (tj. budoucí) uživatele. Jejich vymezení je tedy stanovením znalostní základny cílové komunity, pro kterou má archiv digitální informace uchovávat v nezávisle srozumitelné podobě, a na základě tohoto stanovení shromažďovat a uchovávat potřebné interpretační informace.

### 2.2.1.3 Prostředí archivu OAIS

Norma ISO 14721 vymezuje okolí, ve kterém se archiv OAIS nachází a které určuje jeho vstupy a výstupy, zejména z hlediska informací.

Producent (*producer*) je „úloha vykonávaná osobami nebo klientskými systémy poskytujícími informace určené k uchovávání; může se jednat o další archivy OAIS nebo také o osoby či systémy v daném archivu OAIS“ (10, s. 25).

Koncový uživatel (*consumer*) je „úloha vykonávaná osobami nebo klientskými systémy, které využívají služeb archivu OAIS za účelem nalezení a vlastního zpřístupnění uchovávaných informací; tuto úlohu mohou vykonávat další archivy OAIS nebo též osoby nebo systémy z daného archivu OAIS.“ (10, s. 21).

Management (*management*) je „úloha vykonávaná těmi, kdo určují celková pravidla archivu OAIS jako součást širších pravidel, například v rámci větší organizace“ (10, s. 24).

Vztahy archivu OAIS s producentem nebo koncovým uživatelem vymezují vnější interakce archivu OAIS, vztahy mezi archivem OAIS a managementem jeho vnitřní interakce. Management může například ve vztahu k archivu schvalovat zřizovací listinu, určovat rozsah působnosti, být hlavním zdrojem financování nebo vyhodnocovat výkon. V praxi knihoven je digitální archiv obvykle užší organizační útvar, který je součástí knihovny jakožto konkrétní organizace zřízené za účelem širšího poslání, než je pouze dlouhodobé uchovávání digitálních informací. Tento útvar nemusí být ani formálně vymezen (mohou se na něm podílet zaměstnanci různých oddělení), jako je tomu například ve francouzské národní knihovně. Managementem je vedení knihovny (např. generální ředitel se svými náměstky), které určuje obecná pravidla organizace, která mají zásadní vliv i na výkon archivu OAIS (např. přerozdělení nebo zajištění nových pracovních úvazků). Producent je podle normy ISO 14721 role spočívající v dodávání informací do archivu OAIS. V praxi masové digitalizace knihoven vykonává úlohu producenta, archivu i managementu často jedna a tatáž knihovna.

Z definice normy ISO 14721 vyplývá, že hlavní činností producenta je dodávat informace do archivu. Tato norma však dále vymezuje také některé další dílčí činnosti,

ze kterých vyplývá, že producent je často též samotným producentem informací (a samozřejmě sám výraz „producent“ tuto funkci jasně implikuje). Norma 16363 zavádí synonymické pojmy „poskytovatel (provider)“ a „dodavatel (submitter)“ a definuje je jako „osoba nebo systém dodávající digitální objekt do repozitáře; poskytovatelem může být producent“ (11, s. 17). Motivací této odchylky od normy ISO 14721 je nejspíš reflexe skutečnosti, že v praxi nastávají případy, kdy informace do archivu mohou dodávat subjekty, které nejsou jejich vlastními producenty. Tato úprava je však spíše formální než věcná vzhledem k tomu, že hlavní činností producenta podle normy ISO 14721 není produkovat, ale poskytovat (*provide*) informace, a hlavní činností poskytovatele (nebo dodavatele) v normě 16363 je dodávat (*submit*) informace, což je prakticky totéž (norma ISO 14721 ostatně nazývá dohodu mezi producentem a archivem OAIS „submission agreement“). Rozdíl mezi normou ISO 14721 a normou 16363 tedy v konečném důsledku nespočívá v úpravě definice, ale v pouhé náhradě (v některých případech) zavádějícího termínu „producent“.

Archiv OAIS podle normy ISO 14721 uzavírá s vkladatelem dohodu o dodávání dat (*submission agreement*), což je „dohoda uzavřená mezi archivem OAIS a producentem, která stanovuje datový model a další potřebná nastavení pro relaci dodávání dat (*data submission session*); datový model určuje formát/obsah a logické konstrukty užívané producentem a způsob, jakým jsou reprezentovány na všech dodaných datových nosičích nebo při všech telekomunikačních spojeních“ (10, s. 26). Relace dodávání dat je „jednotlivá dodávka datového nosiče nebo telekomunikační spojení, kterými jsou archivem OAIS poskytována data“ (10, s. 22).

Tato dohoda by podle normy ISO 14721 měla vždy v nějaké podobě existovat, ale nemusí jít vždy o formální podobu (smlouvu). Uváděným příkladem je webarchiv (jako typ archivu OAIS, který uchovává sklizený webový obsah), kde dohoda o dodávání dat nabývá podobu nastavení sklízecího robota (10, s. 36).

S koncovým uživatelem pak archiv OAIS uzavírá dohodu o objednávce (*order agreement*), což je „dohoda mezi archivem a koncovým uživatelem, v níž jsou stanoveny údaje o dodání, například typ datového nosiče a formát dat“ (10, s. 24). Tato dohoda opět nemusí být formální a ustanovení normy lze interpretovat tak, že dohodou o objednávce může být jednoduše to, že uživatel v digitální knihovně (jako součásti archivu OAIS) vyhledá požadovaný dokument.<sup>8</sup> Rozdíl mezi koncovým uživatelem a cílovou komunitou spočívá

---

<sup>8</sup> Zejména viz (10, s. 38).

v tom, že koncový uživatel je jakýkoliv subjekt, který interaguje s archivem OAIS s cílem získání informací (tedy i softwarový systém). Člen cílové komunity je takový koncový uživatel, na základě jehož znalostní základny se udržují informace tak, aby byly srozumitelné samy o sobě.

#### **2.2.1.4 Povinnosti archivu OAIS**

Norma ISO 14721 vymezuje šest obecných závazných povinností, které musí daná organizace plnit, aby mohla provozovat archiv OAIS. Tyto povinnosti odlišují pojem „archiv OAIS“ od jiných užití pojmu „archiv“. Jde o následující povinnosti (10, s. 39):

1. Vyjednávat s producenty informací a přijímat od nich příslušné informace.
2. Získávat možnost s poskytnutými informacemi dostatečně nakládat, aby bylo možné zajistit jejich dlouhodobé uchovávání.
3. Určit, ať již samostatně nebo ve spolupráci s dalšími stranami, které komunity by se měly stát cílovými komunitami, a tudíž by měly být schopny porozumět poskytovaným informacím. Tím je vymezena znalostní základna dané skupiny.
4. Zajistit, aby informace určené k uchovávání byly pro cílovou komunitu srozumitelné samy o sobě. Cílová komunita by měla být schopna informacím porozumět bez využití odborných zdrojů, například bez rady odborníků, kteří informace vytvořili.
5. Dodržovat zdokumentovaná pravidla a postupy, které zajistí, že informace budou chráněny před všemi možnými nepředvídatelnými událostmi (včetně zániku archivu), a zajistit, že informace nebudou nikdy smazány (s výjimkou případu, kdy jejich smazání bude součástí schválené strategie). Nemělo by docházet k žádnému jednorázovému mazání dat.
6. Zpřístupňovat uchovávané informace cílové komunitě a umožňovat šíření informací v podobě kopií původně dodaných datových objektů nebo v takové podobě, aby bylo možné zpětně dohledat (*as traceable to*), ke kterým původně dodaným datovým objektům se vztahují, a to společně s doklady o jejich autenticitě.

První povinnost zahrnuje zejména dohodu o dodávání dat. Druhá povinnost zahrnuje vyřešení otázky práv duševního vlastnictví (pokud archiv nemá právo dodaná data měnit, například je převést do jiného formátu, je jeho činnost *de facto* znemožněna). Třetí povinnost se týká cílové komunity z hlediska jejího vymezení. Čtvrtá povinnost se týká cílové komunity s ohledem na související koncepty nezávislé srozumitelnosti (tj. informace srozumitelné samy

o sobě) a interpretačních informací. Pátá povinnost specificky pojednává o problematice dlouhodobého uchovávání navzdory změnám technologií (datových nosičů, formátů apod.).

Šestá povinnost se týká zpřístupňování informací a specificky zohledňuje otázku autenticity. O tomto aspektu pojednává oddíl páté kapitoly disertační práce zabývající se konceptem autenticity v normě ISO 14721:2012.

### **2.2.2 Informační model**

Informační model přináší typologii informačních objektů, které se vyskytují v průběhu životního cyklu informací. Z nich musíme vyčlenit jako klíčový informační objekt ten, který norma ISO 14721 nazývá informační obsah a který je podle normy hlavním předmětem dlouhodobého uchovávání v archivu OAIS; dále pak zmíněné interpretační informace. Kategorie informací, které slouží jako nezbytné dodatečné informace, jež musejí být uchovávány spolu s informačním obsahem, nazývá norma archivační informace. Ty obsahují pět podkategorií informací. Dalším důležitým prvkem informačního modelu je informační balíček jako logická schránka, která zpravidla obsahuje informační obsah a archivační informace. Důležitým novým prvkem druhého vydání normy ISO 14721 je pojem transformační vlastnost informací.

#### **2.2.2.1 Informační obsah a interpretační informace**

Informační obsah (*content information*) je podle normy ISO 14721 „množina informací, která je původním předmětem uchovávání nebo která obsahuje část těchto informací či všechny tyto informace; informační obsah je informační objekt složený ze svého datového objektu s obsahem a svých interpretačních informací.“ (10, s. 21). První vydání obsahovalo následující definici informačního obsahu: „množina informací, která je původním předmětem uchovávání; informační obsah je informační objekt složený ze svého datového objektu s obsahem a svých interpretačních informací.“ (12, s. 1-8). Původní vydání normy uvádí jako příklad informačního obsahu tabulku s čísly, která reprezentuje teplotní údaje (a která je chápána jako teplotní údaje), s tím, že součástí informačního obsahu již není dokumentace, který by vysvětlovala historii a původ těchto teplotních údajů nebo to, jak se vztahují k jiným pozorováním. Tato změna souvisí s přijetím konceptu, že některé části informačního obsahu nemusejí (nebo nemohou) být zachovány.

Datový objekt s obsahem (dále zkráceně jako objekt CDO) tedy nese informační obsah, pokud je příjemcem spojen s odpovídajícími interpretačními informacemi (které mohou, ale nemusejí být součástí jeho znalostní základny). Je třeba upozornit, že digitálním

objektem podle normy ISO 14721 není nutně soubor – může jím být, stejně jako jím může být například množina souborů. V případě objektu CDO může jít o jeden soubor ve formátu PDF, který reprezentuje digitální knihu, nebo stovku souborů ve formátu JPEG, které dohromady reprezentují sto stran knihy – záleží na konkrétní implementaci. Jednotlivý digitální objekt CDO v archivu je tedy taková podmnožina bitových posloupností z celkové množiny všech bitů uložených na datových nosičích archivu, které jsou potřebné k transformaci bitů do daného konkrétního informačního obsahu.

Interpretační informace člení norma ISO 14721 na tři druhy:

Strukturální interpretační informace (*structure information*) jsou takové „interpretační informace, které udávají, jak jsou další informace složeny; mohou například převádět bitové toky (*bit streams*) na základní typy dat, jako jsou znaky, čísla a pixely, a na seskupení těchto typů dat, jako znakové řetězce a pole“ (10, s. 26). Sémantické interpretační informace (*semantic information*) jsou takové „interpretační informace, které podrobněji popisují význam nesený strukturálními interpretačními informacemi.“ (10, s. 26).

Třetí druh interpretačních informací zavedlo až druhé vydání normy – jde o ostatní interpretační informace (*other representation information*), které jsou definovány jako takové „interpretační informace, které nelze jednoduše zařadit mezi strukturální interpretační informace nebo sémantické interpretační informace; například k porozumění datovému objektu s obsahem mohou být potřeba software, algoritmy, šifrování, psané pokyny atd., přičemž všechny tyto informace budou odpovídat definici interpretačních informací, byť nebude zřejmé, zda se vztahují ke struktuře nebo významu; informace udávající vztah mezi strukturálními interpretačními informacemi a sémantickými interpretačními informacemi nebo popisující software potřebný pro zpracování databázového souboru lze také považovat za ostatní interpretační informace.“ (10, s. 24).

Výše byl uveden příklad interpretačních informací potřebných k převedení fyzického objektu (tištěné knihy) do informačního obsahu (tedy člověku srozumitelného intelektuálního obsahu). Lze říci, že učebnice angličtiny a slovník jsou datové objekty, které reprezentují sémantické interpretační informace. Znalostní základna čtenáře, který si je opatřil, přitom již předtím obsahovala strukturální interpretační informace, jež mu umožňovaly ve fyzickém objektu vidět lineární text (strukturu).

Nás zde zajímá především otázka interpretace digitálních objektů. Lze říci, že problematika interpretačních informací tvoří jádro digitální archivace. Digitální objekt musí být v daný okamžik vždy uložen na nějakém datovém nosiči, který je sice fyzickým



objektem (pevný disk, BluRay Disc apod.), ale data na něm zapsaná člověk vnímat nedokáže, natož je převést do srozumitelné podoby. Z tohoto hlediska je nutné, aby interpretaci digitálního objektu zprostředkovávala počítačová technologie.

V praxi digitální archivace se klade největší důraz na strukturální interpretační informace, zejména na formát souboru. Formát jako jiné označení těchto interpretačních informací pak dále v textu uvádí sama norma ISO 14721 (10, s.66). Jedná se dodatečně vysvětlení významu tohoto pojmu, které přineslo až druhé vydání. Otázce formátu se přikládá v digitální archivaci paměťových institucí nejvyšší důležitost. Formáty patrně nejviditelněji odrážejí problém zastarávání digitálních technologií. Blíže se tomuto problému věnuje oddíl 3.1.

Z hlediska své činnosti musí archiv OAIS shromáždit dostatečné interpretační informace k objektu CDO. Způsoby konkrétního řešení se mohou lišit podle posouzení situace archivem. V případě formátu to může znamenat získat formátovou specifikaci a uložit ji spolu s objektem CDO nebo jen zaznamenat technické informace o formátu, na základě kterých je možno tyto interpretační informace dohledat (tj. předpoklad, že jsou tyto informace běžně dostupné, a tedy samy o sobě srozumitelné pro cílovou komunitu). Tyto technické informace jsou v praxi zaznamenávány do datových objektů, které se označují jako technická metadata. Z informačního modelu OAIS je zřejmé, že i k těmto metadatům (tak jako k jakýmkoliv jiným typům dat), je nutné získat dostatečné interpretační informace. Tento proces podle normy ISO 14721 v digitálním světě znamená rekurzivitě – interpretační informace o technických metadatach budou zaznamenány v dalším digitálním objektu (např. v PDF popisujícím metadatovým standard). Tato rekurzivita podle normy končí tehdy, pokud jsou interpretační informace zaznamenány v podobě fyzického objektu (tedy např. vytištěný standard). V praxi to v současnosti není považováno za větší problém vzhledem k tomu, že metadata jsou zpravidla ukládána v textových formátech (zejména v XML), jejichž interpretace se nepovažuje za problematickou.

Norma ISO 14721 uvádí dva typy specializovaných softwarových nástrojů pro interpretační informace (tyto nástroje samy patří mezi ostatní interpretační informace). Software pro zobrazení interpretačních informací (*representation rendering software*)<sup>9</sup> je software, který umožňuje interpretační informace reprodukovat v podobě srozumitelné lidem. Příkladem je prohlížeč Adobe Acrobat, který dokáže zobrazit formátovou specifikaci

---

<sup>9</sup> Anglický název je zavádějící, měl by být spíše „representation information rendering software“).

uloženou v PDF. Zpřístupňovací software (*access software*) je software, který dokáže prezentovat samotný informační obsah nebo jeho část. Jinými slovy, software pro zobrazení interpretačních informací reprodukuje interpretační informace a slouží jako pomůcka pro uchovávání informačního obsahu, zatímco zpřístupňovací software reprodukuje informační obsah a slouží jako prostředek pro zpřístupňování informačního obsahu cílové komunitě.

Na tomto místě je třeba upozornit, že norma ve své obecnosti nepopisuje odlišnosti užití různých typů informačního obsahu. Archiv OAIS musí uchovávat objekt CDO spolu se zvolenou množinou interpretačních informací (které bude uchovávat v podobě dalšího digitálního objektu, zpravidla nazývaného metadata). Některé typy informačního obsahu mohou být určeny k reprodukci lidským uživatelům, jako tomu bývá v případě dokumentů digitálního dědictví (např. zobrazení v digitální knihovně), zatímco jiné nemusí být vůbec určeny pro vnímání člověkem, ale pro zpracování jinými systémy (to je příklad rozsáhlých datových sad získaných z pozorování, které mohou být reprezentovány v podobě tabulek, obsahovat numerické údaje a sloužit jako zdroj pro automatizované analýzy).

Za jeden ze způsobů, jak řešit otázku interpretačních informací, bylo určeno vytvoření mezinárodních registrů interpretačních informací (zejména pro formáty). Otázce těchto registrů a bližšímu přiblížení problematiky formátů se věnuje Kapitola 3.

Závěrem tohoto shrnutí je potřeba zopakovat důraz na rozlišování interpretačních informací na jedné straně a metadat, které je reprezentují, na straně druhé. Metadata jsou datové objekty a potřebují rovněž své interpretační informace. Rekurzivitu v praxi ukončuje užití souborů v textových formátech typu XML, jejichž interpretace se z hlediska digitální archivace považuje (vzhledem k jednoduchosti souborového formátu) za relativně bezproblémovou.

#### **2.2.2.2 Archivační informace**

Informační model OAIS definuje archivační informace (*preservation description information*) jako „informace, které jsou nezbytné k adekvátnímu uchovávání jednotlivého informačního obsahu“ (11, s. 25). Jde o kategorii informací, která obsahuje pět podkategorií.

Identifikační informace (*reference information*) jsou „informace, která jsou využívány jako identifikátor informačního obsahu“ (10, s. 25). Tyto informace mohou také zahrnovat „identifikátory, které vnějším systémům umožňují jednoznačně odkazovat na konkrétní informační obsah“ a udávat a popisovat způsob jejich přidělování (10, s. 74). Uváděným příkladem z oblasti knihoven je perzistentní identifikátor (konkrétním příkladem pak ISBN)

a bibliografický popis. Tyto informace tedy mohou zahrnovat i globální perzistentní identifikátory, na základě nichž mohou uživatelé vyhledávat konkrétní informační obsah.

Provenienční informace (*provenance information*) jsou „informace, které dokumentují historii informačního obsahu; tyto informace vypovídají o původu nebo zdroji informačního obsahu, o veškerých změnách, které mohly od doby jeho vzniku nastat, a o tom, kdo o něj od doby jeho vzniku pečoval“ (10, s. 25). Jak dále norma uvádí, archiv nese odpovědnost za vytváření a uchovávání těchto informací až od okamžiku jejich příjmu do archivu; provenienční informace z dřívější doby by měl poskytnout vkladatel. Uváděnými příklady z oblasti knihoven jsou metadata o procesu uchovávání (ukazatele k předchozím verzím jednotky, historie změn); metadata o procesu digitalizace. Součástí provenienčních informací je popis vlastností informací, který je popisem vlastností informačního obsahu, jež by měly být dlouhodobě uchovávány (a který také může zahrnovat vyjádření toho, co nemusí být zachováno) (10, s. 50). Tento popis může dodat producent, nebo jej může vytvořit archiv. Provenienční informace, jako jediný z dílčích typů archivačních informací, dává norma do souvislosti s autenticitou; o tomto aspektu pojednáme později.

Informace o neporušenosti (*fixity information*) jsou „informace, které udávají, jak je zajištěno, aby objekt s informačním obsahem nebyl nezdokumentovaným způsobem změněn“ (10, s. 22). Tyto informace „fungují jako obal nebo ochranný štít, který chrání informační obsah“ (10, s. 34). Příkladem je digitální otisk. Dalším příkladem je indikátor autenticity (*authenticity indicator*), tento pojem však v normě není nijak definován a vyskytuje se jen na jediném místě (jako příklad) (10, s. 75), proto se jím ve výzkumné části nebudeme zabývat. Norma dále uvádí, že do této dílčí kategorie nepatří „mechanismy zajišťující integritu (integrity), které jsou poskytovány službami archivu OAIS“ a „ochrana před chybami poskytovaná datovými nosiči“ (10, s. 74). Pojem integrity není v normě definován, nicméně na jiném místě norma popisuje funkční prvek archivu OAIS nazvaný „Služba integrity dat“, který „zajišťuje, aby data nebyla nepovoleným způsobem změněna nebo zničena. Tato služba se vztahuje k datům v trvalých úložištích dat a k datům v přenášených zprávách.“ (10, s. 49).

Kontextuální informace (*context information*) jsou „informace, které dokládají vztah informačního obsahu k jeho okolí; patří mezi ně důvod vytvoření informačního obsahu a jeho vztah k dalším objektům s informačním obsahem.“ (10, s. 21). Provenienční informace lze podle normy považovat za zvláštní typ kontextuálních informací (10, s. 49), což ukazuje na menší nekonzistenci v rozdělení kategorie archivačních informací na pět podkategorií.

Informace o přístupových právech (*access rights information*) jsou „informace, které udávají omezení týkající se přístupu k informačnímu obsahu, a to včetně právního rámce, licenčních podmínek a řízení přístupu“ (10, s. 19). Tato dílčí kategorie byla přidána druhým vydání normy ISO 14721.

### 2.2.2.3 Informační balíček

Informační balíček je „logická schránka, která může obsahovat informační obsah a archivační informace; k tomuto informačnímu balíčku jsou připojeny informace o zabalení (*packaging information*), které vymezují a určují informační obsah, a informace o popisu balíčku (*package description*), které usnadňují vyhledání informačního obsahu“ (10, s. 23).

Norma dále odlišuje tři typy informačních balíčků:

- Archivní informační balíček (*archival information package*) je „informační balíček, který je složen z informačního obsahu a přidružených archivačních informací a je uchováván v archivu OAIS“ (10, s. 20). Dále bude nazýván jen jako balíček AIP.
- Vstupní informační balíček (*submission information package*) je „informační balíček, který dodává producent do archivu OAIS tak, aby mohl být využit při sestavení nebo aktualizaci jednoho nebo více AIP a/nebo přidružených popisných informací (*descriptive information*)“ (10, s. 26). Dále bude nazýván jako balíček SIP.
- Výstupní informační balíček (*dissemination information package*) je „informační balíček odvozený z jednoho nebo více balíčků AIP a zaslaný archivem OAIS koncovému uživateli jako odpověď na jeho požadavek vůči tomuto archivu (10, s. 22). Dále bude nazýván jako balíček DIP.

Klíčový informační balíček z hlediska archivu OAIS je balíček AIP. Ten musí obsahovat informační obsah a přidružené archivační informace. Definice nadřazené kategorie (tj. informačního balíčku) uvádí, že tyto dvě složky může obsahovat. To znamená, že při dodávání balíčků SIP do archivu nebo vydávání balíčků DIP koncovým uživatelům nemusí každý jednotlivý balíček SIP nebo DIP vždy obsahovat informační obsah a archivační informace. Do archivu mohou být například dodávány odděleny balíčky SIP s informačním obsahem a balíčky SIP s archivačními informacemi. Podobně mohou být archivem vydávány, v závislosti na požadavcích koncových uživatelů například, jen balíčky DIP obsahující informační obsah, ale již nikoliv archivační informace.

Informace o zabalení (*packaging information*) jsou „informace, které slouží k propojení a popisu součástí informačního balíčku“ (10, s. 24). Popis balíčku (*package*

*description*) jsou „informace určené pomůckám pro zpřístupnění“ (10, s. 24). Těmito pomůckami norma míní „softwarový program nebo dokument, který koncovým uživatelům umožňuje najít, analyzovat, objednat nebo získat informace z archivu OAIS“ (10, s. 19). Popisné informace (*descriptive information*) jsou „množina informací, která je složena především z popisů balíčků a je poskytována správě dat za účelem podpory koncových uživatelů při objednávání a získávání informačních jednotek z archivu OAIS“ (10, s. 22).

Hlavním účelem informací o zabalení je vymezit, které části balíčku AIP jsou informační obsah a které archivační informace. To znamená popsat jednak to, které soubory v balíčku tvoří objekt CDO, jež reprezentuje informační obsah (např. pět souborů ve formátu JPEG obsahujících články) a které technická metadata reprezentující interpretační informace (např. jeden soubor v XML popisující formát JPEG). Objekt CDO a technická metadata dohromady tvoří informační obsah. Dále je třeba popsat, jaké soubory jsou archivační metadata reprezentující archivační informace (např. druhý soubor v XML obsahující bibliografický popis). V popisu balíčků a v popisných informacích se pak podle normy obvykle opakují identifikační informace. Pokud budeme uvažovat digitální knihu v PDF, pak informační obsah může tvořit jeden soubor ve formátu PDF (objekt CDO) a jeden soubor ve formátu XML obsahující technická metadata (např. informace o verzi a typu, příkladem je PDF/A-2u); archivační informace jeden soubor v XML s archivačními metadaty (blíže o metadatach bude pojednávat Kapitola 3), přičemž jejich součástí bude identifikátor ISBN (který bude současně obsažen i v samotném informačním obsahu, jelikož je vydavatelskou praxí uvádět jej přímo v knize) a tento identifikátor bude také obsažen v popisu balíčku a popisných informacích. Na základě tohoto identifikátoru pak čtenář jakožto koncový uživatel může knihu vyhledat v archivu.

Předmětem dlouhodobého uchovávání je informační obsah (např. konkrétní kniha v PDF) i přidružené archivační informace uložené v balíčku AIP spolu s informačním obsahem. Archiv vytváří balíček AIP z balíčku SIP (nebo z více balíčků SIP), který mu dodá vkladatel. Na základě dohody o dodávání dat oba subjekty (archiv a vkladatel) definují datový model, tedy podobu dodávaných informací, jehož součástí by měl být datový slovník (popis všech typů dat, což zahrnuje i metadata). Osvědčeným postupem v komunitě digitální archivace je dohoda o užití takových formátů objektu CDO, které podporují dlouhodobé uchovávání, a užití mezinárodně rozšířených metadatových schémat. Kontext, kterému se věnuje tato disertační práce, je situace, kdy vkladatel je současně i producentem a úloha archivu i vkladatele je vykonávána stejnou organizací, takže je celý životní cyklus lze jednotně naplánovat a řídit podle osvědčených postupů. Ovšem i v případě, kdy vkladatel

nemůže ovlivnit podobu formátu objektu CDO (např. není producent), je technicky v jeho moci, alespoň v teoretické rovině, dodat v požadované podobě alespoň metadata.

### 2.2.3 Funkční model

Funkční model OAIS vymezuje šest základních funkčních celků (*functional entities*) archivu OAIS. Funkční celek je konceptuální model, představující kategorii činností (funkcí a služeb), které musejí být archivem vykonávány, aby byl archivem OAIS. Výkon těchto činností pro digitální archivaci zajišťují počítačové technologie stejně jako lidští pracovníci (nejen jako obsluha, ale i jako tvůrci návrhu softwaru apod.). Těmito funkčními celky jsou: Příjem, Archivní úložiště, Správa dat, Administrace, Plánování uchovávání a Zpřístupnění (10, s. 45). Sedmým funkčním celkem jsou základní služby (*common services*), jimiž jsou „podpůrné služby nezbytné pro provoz archivu OAIS“ (10, s. 20). Ty nejsou znázorněny v obrázku, neboť souvisí se všemi šesti základními funkčními celky. Funkční celek je detailněji popsán jako výčet funkčních prvků (*functions*), což jsou skupiny dílčích činností (podkategorie činností) vykonávaných v rámci jednoho funkčního celku. V následujících podřazených oddílech je postupně uveden přehled všech funkčních prvků, vybrány jsou však pouze některé činnosti, aspekty a vazby, které mohou být relevantní pro rámec této disertační práce.

#### 2.2.3.1 Funkční celek Příjem

Funkční celek Příjem (*Ingest functional entity*) poskytuje služby a funkce vztahující se k balíčkům SIP a k přípravě informačního obsahu na uložení a správu v rámci archivu. Balíček SIP může být dodáván do archivu nejen vkladatelem, ale také samotným archivem (za účelem aktualizace).

Funkční celek Příjem zahrnuje tyto funkční prvky (10, s. 49-51):

- **Příjem vstupních dat**
  - dočasné úložiště pro uložení balíčku SIP
  - potvrzení o přijetí balíčku SIP vydávané vkladateli
  - požadavek na opětovné dodání balíčku SIP (z důvodu opravy)
- **Kontrola kvality**
  - provedení kontroly neporušenosti při přesunu balíčku SIP na dočasné úložiště
- **Vytváření balíčku AIP**
  - převod balíčků SIP do balíčku AIP v souladu se standardy archivu pro formáty dat a dokumentaci

- normalizace (formátová konverze do preferovaného formátu)
  - shromáždění potřebných interpretačních informací
- zasílání balíčků SIP/AIP funkčnímu celku Administrace k provedení kontroly
  - opakovaný převod na základě požadavků vzešlých z této kontroly
- **Vytváření popisných informací**
  - vyjímání popisných informací z balíčků AIP
  - shromažďování popisných informací z dalších zdrojů
- **Sladění aktualizací**
  - přesun balíčků AIP do Archivního úložiště
  - získání informace o místě uložení balíčku AIP
  - přesun kompletních popisných informací (vč. informace o místě uložení balíčku AIP) do databáze Správy dat

### 2.2.3.2 Funkční celek Archivní úložiště

Funkční celek Archivní úložiště (*Archival storage functional Entity*) poskytuje služby a funkce vztahující se k ukládání, údržbě a získávání balíčků AIP.

Funkční celek Archivní úložiště zahrnuje tyto funkční prvky (10, s. 51-53):

- **Příjem balíčků AIP**
  - příjem balíčků z funkčního celku Příjmu
  - přesun balíčků AIP do trvalého úložiště
  - udání očekávané četnosti využití balíčků AIP
- **Správa struktury úložiště**
  - výběr vhodného nosiče
  - sledování chybových záznamů
  - provozní statistiky pro funkční celek Administrace:
    - soupis datových nosičů
    - dostupné úložné kapacity
    - statistiky užití
- **Nahrazení datových nosičů**
  - přesuny balíčků AIP na nové datové nosiče v průběhu času<sup>10</sup>
- **Kontrola chyb**
  - kontrola, zda při přesunu na jiný datový nosič nebo v průběhu uložení na datových nosičích nedošlo k poškození balíčku AIP
  - schopnost upozorňovat na chyby hardwaru a softwaru
- **Obnova po havárii**
  - mechanismy pro vytváření záložních kopií celé sbírky balíčků AIP na fyzicky oddělená zařízení
- **Poskytování balíčků AIP**
  - přijetí požadavku na získání balíčku AIP z funkčního celku Zpřístupnění
  - vydání balíčků AIP na požadovaných datových nosičích nebo do dočasného úložiště funkčního celku Zpřístupnění

---

<sup>10</sup> Při tomto přesunu nesmí být změněn ani informační obsah, ani archivační informace. Změnit lze pouze informace o zabalení (je-li to nutné), pokud to ovšem nezpůsobí ztrátu informací. Příkladem takového přesunu je přesun na nové datové nosiče v novém operačním systému a novém souborovém systému (za předpokladu, že informační obsah a archivační informace nejsou závislé na souborových systémech). Norma tímto chce odlišit relativně rutinní činnost výměny datových nosičů od složitějších případů, jako je formátová konverze.



### 2.2.3.3 Funkční celek Správa dat

Funkční celek Správa dat (*Data management functional entity*) poskytuje služby a funkce pro vložení, údržbu a zpřístupnění jak popisných informací, tak administrativních dat využívaných při správě archivu.

Funkční celek Správa dat zahrnuje tyto funkční prvky (10, s. 54-55):

- **Správa databáze**
  - Udržování integrity databáze
  - Dostatečné prostředky pro ukládání:
    - popisných informací
    - systémových informací<sup>11</sup>
- **Provádění dotazů**
  - přijetí požadavku z funkčního Zpřístupnění a odpověď na něj
- **Vytváření přehledů**
  - vytváření přehledů z funkčních celků Příjmu, Zpřístupnění nebo Administraci:
    - seznamů archivních jednotek podle kategorie
    - statistik o přístupech uživatelů
  - poskytování popisných informací k balíčkům AIP
- **Příjem aktualizací databáze**
  - Přidává, upravuje nebo maže informace z trvalého úložiště funkčního celku Správa dat, přičemž mezi jeden z hlavních zdrojů aktualizací patří popisné informace k novým balíčkům AIP
  - Podává pravidelné přehledy o stavu aktualizací databáze funkčnímu celku Administrace

Funkční celek Správa dat se týká řízení popisných informací (pro účely správy a vyhledávání v archivu, zejména balíčků AIP) a administrativních dat, nikoliv řízení změn dat obsažených v balíčcích AIP. Možná by bylo vhodnější označit tento celek jako „Správa metadat“ nebo „Správa systémových dat“.

---

<sup>11</sup> Tj. informací užívaných pro provoz archivu.

#### 2.2.3.4 Funkční celek Administrace

Funkční celek Administrace (*Administration functional entity*) poskytuje služby a funkce sloužící k provozu archivního systému jako celku.

Funkční celek Administrace zahrnuje tyto funkční prvky (10, s. 55-57):

- **Vyjednávání dohody o dodávání dat**
- **Správa nastavení systému**
  - sledování fungování archivu a řízení změn nastavení
  - zajištění možnosti zpětné zjištělnosti všech minulých nastavení
- **Aktualizace archivovaných informací**
- **Řízení fyzického přístupu**
  - prostředky, které zamezují nebo kontrolují fyzický přístup (dveře, zámky, ochranka apod.)
- **Stanovování standardů a pravidel**
- **Kontrola vstupních dat**
- **Spouštění požadavků**
  - Požadavek na distribuci balíčků AIP zasílaný do funkčního celku Zpřístupnění
- **Služba zákazníkům**
  - Řízení účtů koncových uživatelů (zejména pro platící zákazníky)
  - Odpovídání na požadavky na získání obecných informací
  - Shromažďování zpětné vazby týkající se služeb pro zpřístupnění
  - Vytváření a zpřístupňování přehledů připomínek

Funkční celek Administrace je hlavní řídicí část archivu OAIS a některé výše uvedené funkční prvky je nutno blíže popsat, aby byly srozumitelné.

Funkční prvek Vyjednávání dohody o dodávání dat zahrnuje vyjednávání dohody o dodávání dat s vkladateli a o harmonogramu tohoto dodávání. Formáty a postupy by měly být podrobně zdokumentovány v pravidlech archivu pro formáty a postupy pro dodávání dat. Podobu této dokumentace norma ISO 14721 podrobněji nepopisuje, ale odpovídá konceptu pravidel pro uchovávání (*preservation policies*), který je uveden v normě ISO 16363 jako typ dokumentu, který musí archiv mít sepsaný (11, s. 17). Ten pak tvoří součást dokumentace archivu, na základě které se posuzuje jeho důvěryhodnost. Zejména je důležité, aby pravidla pro uchovávání obsahovala seznam archivem preferovaných formátů. Na základě dohody o dodávání dat by pak archiv měl s vkladatelem sjednat, zda mu vkladatel bude dodávat

objekty CDO v těchto preferovaných formátech, nebo jestli bude archiv provádět jejich normalizaci, tedy převod do preferovaných formátů (viz funkční prvek Vytváření balíčku AIP funkčního celku Příjem). Šablony balíčků SIP/AIP (stanovenou podobu dodávání) přijímá od funkčního celku Plánování uchovávání.

Funkční prvek Aktualizace archivovaných informací je klíčovým prvkem dlouhodobého uchovávání. Představuje realizaci opatření z důvodu změn technologického prostředí. Funkční celek Administrace řídí tento proces tak, že iniciuje vydání balíčků AIP (jako balíčku DIP), provede jejich aktualizaci (např. formátovou konverzi) a jako balíčky SIP je dodá do funkčního celku Příjmu tak, jak to činí vkladatel. Dalším proces je téměř stejný s tím rozdílem, že v tomto případě zdrojem dat není vkladatel, ale sám archiv. Blíže k problematice aktualizace archivovaných informací viz oddíl Přístupy k uchovávání (2.2.4).

Funkční prvek Řízení fyzického přístupu ukazuje, že archiv není jen otázkou počítačových technologií a lidské činnosti, ale také fyzického prostoru. Otázky přidělení tohoto prostoru (budovy, sálů apod.) nejsou předmětem činnosti archivu OAIS, ale jeho managementu, v knihovnách typicky vedení knihovny nebo jejího zřizovatele.

Funkční prvek Stanovování standardů a pravidel v první řadě zajišťuje součinnost s managementem archivu OAIS. Pro management zajišťuje pravidelné přehledy o činnosti archivu OAIS a od managementu přijímá organizační pravidla (zřizovací listina, rozsah působnosti, směrnice pro využití zdrojů). Od funkčního prvku Plánování uchovávání přijímá nové standardy a doporučení pro archivované informace a pravidelné zprávy, které vyhodnocují aktuální rizika (mj. zastarávání technologií). Vydává standardy pro formáty, dokumentaci a postupy pro proces příjmu. Schvaluje standardy a cíle pro aktualizaci archivovaných informací (dodané funkčním celkem Plánování uchovávání). Vytváří pravidla pro správu úložiště, kterými se řídí funkční celek Archivní úložiště (zejména pravidla pro nahrazování datových nosičů a obnovu po havárii), a pravidla pro bezpečnost obsahu (pravidla pro řízení fyzického přístupu a postupy pro řešení chyb). Provádí kontrolu vstupních dat (která mu zasílá funkční celek Příjem), zejména soulad balíčků SIP nebo AIP s požadavky stanovenými dohodou o dodávání dat. Dále také rozhoduje o tom, zda bude muset vkladatel opětovně dodat chybný balíček, a přijímá (od funkčního celku Plánování uchovávání) posudky, zda byly správně užity šablony balíčků SIP/AIP. Také posuzuje srozumitelnost dat pro cílovou komunitu a kontroluje plnění cílů aktualizace při vytváření nové verze balíčku AIP.

### 2.2.3.5 Funkční celek Plánování uchovávání

Funkční celek Plánování uchovávání (*Preservation planning functional entity*) poskytuje služby a funkce pro monitorování okolí archivu OAIS a doporučení a plány uchovávání k zajištění dlouhodobé přístupnosti a srozumitelnosti informací uložených v archivu OAIS pro cílovou komunitu, a to i v případě zastarání původního počítačového prostředí. Jde o tu část archivu OAIS, která zajišťuje znalostní základnu týkající se způsobu řešení problematiky digitální archivace.

Funkční celek Plánování uchovávání zahrnuje tyto funkční prvky (10, s. 58-60):

- **Sledování cílové komunity**
  - Sledování změn v požadavcích na služby a v dostupných technologiích pro zpřístupnění (ve spolupráci s koncovými uživateli a vkladateli), například požadavky na formát zpřístupňovaných dat
  - Způsoby sledování jsou průzkumy, pravidelné formální hodnocení, skupinové diskuze
- **Sledování technologií**
  - Sledování nových technologií, standardů a platforem (hardwaru, softwaru)
    - Zjišťování, co může způsobit zastarání a zabránit přístupu k informačnímu obsahu
  - Vývoj prototypů pro lepší vyhodnocování nových technologií
- **Tvorba strategií a standardů pro uchovávání**
  - Tvorba strategií a standardů (pro funkční celek Administrativa)
  - Doporučování strategií a standardů
  - Hodnocení rizik
  - Doporučení na rozvoj systému a aktualizace balíčků AIP
- **Tvorba šablon balíčků a plánů migrací**
  - Vytváření nových šablon balíčků SIP a AIP a podrobných plánů migrací
  - Navrhování přizpůsobení a posouzení balíčků AIP/SIP v souvislosti s použitím šablon

### 2.2.3.6 Funkční celek Zpřístupnění

Funkční celek Zpřístupnění (Access Functional Entity) poskytuje služby a funkce, které jednak slouží k podpoře koncových uživatelů (při zjišťování existence, popisu, lokace a přístupnosti archivovaných informací), jednak těmto uživatelům umožňují požadovat a získávat informační produkty.

Funkční celek Zpřístupnění zahrnuje tyto funkční prvky (10, s. 60-61):

- **Sladění činností pro zpřístupnění**
  - Poskytování jednoho nebo více rozhraní pro koncové uživatele pro přístup k archivovaným informacím
  - Zaznamenávání požadavků koncových uživatelů:
    - na zpracování dotazu
    - na získání přehledů
    - objednávek
  - Asistence koncovým uživatelům
- **Vytváření balíčku DIP**
  - Přesun balíčku AIP z funkčního celku Archivního úložiště do dočasného úložiště
  - Provedení případných úpravy balíčků DIP v dočasném úložišti (formátová konverze, odstranění osobních údajů apod.)
- **Dodání odpovědi**
  - Dodání balíčků DIP
  - Dodání odpovědi na dotazy
  - Dodání přehledů
  - Poskytnutí asistence

## 2.2.4 Přístupy k uchovávání

Norma ISO 14721 popisuje dvě hlavní kategorie opatření pro uchovávání informačního obsahu vykonávaná v archivu OAIS nad balíčkem AIP jako digitální migraci a emulaci. Emulaci se blíže věnovat nebudeme (a norma se jí také detailněji nezabývá).

Digitální migrace (*digital migration*) je podle normy „přesun (*transfer*) digitálních informací v rámci archivu OAIS se záměrem tyto informace uchovat.“ (10, s. 22). Podle normy tento přesun od jiných typů přesunů odlišují následující tři atributy (10, s. 22):

- je zaměřen na uchování celého informačního obsahu,
- nová podoba informací v archivu nahrazuje podobu předchozí a
- archiv OAIS řídí všechny stránky přesunu a nese za ně plnou odpovědnost.

Norma ISO 14721 rozlišuje čtyři typy digitálních migrací. Renovace (*refreshment*) a replikace (*replication*) jsou migrace, při kterých nedochází ke změně datových objektů (tedy bitových posloupností), balíčkovací migrace (*repackaging*) a transformace (*transformation*) jsou migrace, při nichž ke změně datových objektů dochází. Definice těchto typů jsou následující (10, s. 103-104):

**Renovace:** Digitální migrace, při níž je jedna instance datového nosiče, která obsahuje balíček AIP, více balíčků AIP nebo části balíčků AIP, nahrazena jinou instancí datového nosiče stejného typu, a to zkopírováním bitů na datový nosič využitý k umístění balíčků AIP a ke správě a přístupu k datovému nosiči. Díky tomu dokáže stávající mapovací infrastruktura archivního úložiště beze změny stále nalézat balíček AIP a přistupovat k němu.

**Replikace:** Digitální migrace, při níž nedochází k žádným změnám balíčkovacích informací, informačního obsahu ani archivačních informací. Bity nesoucí tyto informační objekty jsou při přesunu na novou instanci stejného nebo nového typu datového nosiče zachovány. Renovace je také replikací. Replikace však může vyžadovat změny mapovací infrastruktury archivního úložiště.

**Balíčkovací migrace:** Digitální migrace, při níž dochází k změně bitů balíčkovacích informací.

**Transformace:** Digitální migrace, při níž dochází k změnám bitů informačního obsahu nebo archivačních informací, přičemž je současně vyvinuta snaha uchovat informační obsah v úplnosti.

Renovaci a replikaci se nebudeme dále věnovat. V praxi se souhrnně označuje jako **bitová ochrana** a znamená zkopírování dat na jiný datový nosič, rozdíl spočívá v tom, že replikace může vyžadovat úpravu mapovací infrastruktury (tato změna postihuje hardware a software, nikoliv data). Norma ISO 14721 tento rozdíl nepříliš srozumitelně vysvětluje, ale příklad z běžné praxe pomůže lepšímu porozumění. Zkopírování dat z jednoho nosiče CD na jiný je příkladem renovace. Příkladem replikace je zkopírování dat z deseti nosičů CD na pevný disk počítače.

Žádná z těchto dvou typů migrací nepředstavuje větší intelektuální problém, riziko spočívá především v nedostatku financí a nedostatečné kontroly stavu datových nosičů. Archiv musí mít nastaveny mechanismy pro monitorování stavu datových nosičů a plány pro migraci v případě, že je zaznamenáno riziko jejich degradace nebo zastarávání (tj. konec jejich hardwarové podpory). Rizikem samozřejmě může být i sama volba datových nosičů, které jsou zcela nevhodné svou povahou (názorným příkladem z minulosti je disketa, která nikdy nebyla bezpečným nosičem) nebo náročností údržby (např. z hlediska nedostatku finančních prostředků).

Ani balíčkovací migrace nepředstavuje větší riziko. Jedná se o změnu informací o zabalení. Příkladem může být jiné uspořádání v rámci adresáře. Riziko, které přináší balíčkovací migrace, spočívá spíše v neudržení odlišení informačního obsahu a archivačních informací.

Transformace je nejdůležitějším a nejsložitějším typem digitální migrace. Pouze transformace podle normy zakládá novou verzi balíčku AIP. To znamená, že verze balíčku AIP, která prošla renovací, replikací nebo balíčkovací migrací, zůstává nezměněna.

Původně uložený balíček AIP (tj. informační balíček vytvořený z balíčku SIP dodaného vkladatelem) má být považován za první verzi balíčku AIP a norma jej označuje jako originál, resp. původní balíček AIP (*original AIP*). Tento původní balíček „může být udržován pro ověření uchovávání informací“ (10, s. 105). Norma takovýto postup (tj. uchovávat první verzi balíčku AIP i tehdy, když je vytvořena novější verze) tedy nepředepisuje, pouze uvádí jako možnost.

Transformaci norma dále dělí na dva dílčí typy. Vratná transformace (*reversible transformation*) je taková „transformace, při kterém nová reprezentace stanovuje množinu (nebo podmnožinu) výsledných objektů, které jsou rovnocenné výsledným objektům stanoveným původní reprezentací; to znamená, že novou množinu objektů je možné k původní reprezentaci a její množině základních objektů přiřadit jedna ku jedné“ (10, s. 26).

Uváděným příkladem vratné transformace je migrace textového souboru (obsahujícího písmena anglické abecedy) v kódování ASCII do kódování UNICODE UTF-16. Tento proces lze bezztrátově vrátit (10, s. 106).

Nevratná transformace (*non-reversible transformation*) je taková „transformace, u které nemůže být zaručeno, že se jedná o vratný převod“ (10, s. 24). Uváděný příklad může být nesrozumitelný, proto uveďme jednoduchý příklad z praxe – formátovou konverzi z formátu TIFF (nekomprimovaná verze) do formátu JPEG (formát JPEG nepodporuje jinou než matematicky ztrátovou kompresi).

Norma systematicky nepopisuje rozdíly mezi transformací, která mění informační obsah, a transformací, která mění archivační informace. Její koncepty však lze specifikovat následujícím způsobem. V případě transformace informačního obsahu mohou nastat dvě základní varianty. Zaprvé může být potřeba změnit pouze interpretační informace o formátu. Například může dojít k tomu, že archivem doporučovaný nástroj pro zobrazení souboru ve formátu PDF zastará a nebude adekvátně reprodukovat formát PDF. Pak bude muset archiv provést průzkum a vybrat jiný nástroj, a informace o tomto nově doporučovaném formátu budou zdokumentovány a uchovávány spolu s objektem CDO. Tato transformace tedy zahrnuje pouze interpretační informace. Za druhé může nastat situace, kdy dochází ke změně bitů objektu CDO, typicky v případě formátové konverze. Pokud půjde o konverzi z formátu TIFF (nekomprimovaná verze) do formátu JP2 (matematicky bezztrátová komprese), půjde o vratnou transformaci, nicméně nová reprezentace (objekt CDO ve formátu JP2) bude vyžadovat jiné interpretační informace, neboť se jedná o jiný formát.

Nevratná transformace je podle normy rizikem pro zachování autenticity. O tomto problému v souvislosti s transformací objektu CDO pojednáme v oddíle 5.3. V případě, že jde o transformaci archivačních informací, je možné učinit zobecnění, že v tomto případě by nikdy nemělo jít o nevratnou transformaci, což je dáno tím, že jde o převod metadat z jednoho schématu do nového.

Jinou situací je proces obohacování archivačních metadat. V případě, že se mění objekt CDO, pak se vždy také mění archivační informace v tom smyslu, že se doplní o záznamy těchto změn (zejména jde o provenienční informace). Proto bude v průběhu uchovávání objem archivačních informací narůstat. Rovněž může dojít k tomu, že se v průběhu uchovávání obohatí archivační informace o nové informace (např. další perzistentní identifikátor).



## 2.3 Norma ISO 16363

Některé koncepty normy ISO 16363 již byly představeny v oddíle 2.2 popisujícím normu ISO 14721, jako dokreslení nebo komentář. V tomto úvodním přehledu bude norma ISO 16363 představena pouze stručně v bodech, které jsou relevantní pro tuto disertační práci.

Norma ISO 16363 slouží pro specifikaci konkrétnějších ukazatelů, na základě kterých lze více či méně objektivně doložit, že archiv plní činnosti stanovené normou ISO 14721, a tedy jej lze označit za důvěryhodný digitální archiv (norma ISO 16363 se vztahuje pouze na digitální archiv, tedy dlouhodobé uchovávání pouze digitálních informací).

Důvěryhodností archivu se zde konkrétněji rozumí, že archiv působí prokazatelně v souladu se svými cíli, to znamená, že jeho důvěryhodnost lze doložit navenek (jeho financovatelům, vkladatelům, cílové komunitě i dalším zainteresovaným stranám). Základní premisou této normy je, že musí existovat doklady, podle kterých lze externím auditem plnění požadavků dlouhodobého uchovávání doložit. Doklady přirozeně musejí odpovídat skutečnosti a prokázat, že úložiště má své procesy vhodně nastaveny, řídí se předem zdokumentovanými postupy a dokumentuje všechny podstatné události.

Ukazatele důvěryhodnosti jsou v normě ISO 16363 řazeny do tří základních kategorií: organizační infrastruktura (např. trvalé financování, dostatečné personální zajištění, jasné vymezení rolí a odpovědností, dlouhodobé plánování), správa digitálního obsahu (např. pravidelné sledování vývoje technologií, dodržování standardů, dokumentace všech postupů) a technická infrastruktura a řízení bezpečnostních rizik (např. nepřetržitá údržba a obnova hardwaru a softwaru).

Při prokazování důvěryhodnosti hraje roli celá škála různorodých dokladů: například metadata k digitálním objektům (ze kterých lze např. ověřit historii změny), logy aplikací (umožňují mj. provést kontrolu systémových operací) nebo finanční dokumenty obsahující rozpočtové výdaje (jako doklady o zajištění financování).

Vlastní proces certifikace podle ISO 16363 musí provádět vnější, nezávislý auditor, který splňuje požadavky normy ISO 16919:2014. Certifikace podle normy ISO 16363 je však časově náročný a nákladný proces. Je proto záhodno, aby si instituce nejprve provedly vlastní sebeposouzení podle normy ISO 16363 a na základě něho se nejprve pokusily odstranit všechny zjištěné nedostatky před tím, než podají žádost o certifikaci nezávislému auditorovi. V oblasti řízení bezpečnostních rizik v praxi pak platí, že je potřeba mít nejprve provedenu certifikaci podle normy ISO 27001, a poté usilovat o certifikaci podle normy 16363.

Klíčovou oblastí dokladů je dokumentace. Níže je uveden model dokumentace podle normy ISO 16363 (11, s. 16-17). Tento model vymezuje, co vše má dokumentace obsahovat. V praxi mohou být modelové dokumenty například rozděleny do více dokumentů.

Nejvyšším dokumentem je **mandát archivu** (*repository mission statement*). Jde o písemné prohlášení, které schválil management archivu a které popisuje závazek organizace pečovat o digitální objekty, které má ve své správě. Paměťové instituce zpravidla mívají zřizovací listiny, ne vždy však tyto podrobněji vymezují úkoly digitální archivace a zpravidla tyto instituce vykonávají více činností než je provoz digitálního archivu. Například Národní knihovna ČR má ve své nové zřizovací listině (z roku 2011) závazné pro celou organizaci tento závazek uveden ve velmi obecné rovině.<sup>12</sup> V praxi je proto potřeba formálně vydat dokument popisující, co digitální archivací myslí (základem může být uvedení závazku řídit se normami ISO 14721 a ISO 16363 a specifikace konkrétnějších úkolů) a jaký je rozsah jeho poslání (například od jakých všech vkladatelů bude archiv přijímat informační obsah k dlouhodobému uchovávání).

**Strategický plán uchovávání** (*preservation strategic plan*) je písemné prohlášení, které schválil management archivu a které uvádí cíle a úkoly nutné k dosažení části poslání úložiště vztahující se k uchovávání; strategické plány uchovávání mohou být složeny z dlouhodobých a krátkodobých plánů.

**Pravidla pro uchovávání** (*preservation policy*) je písemné prohlášení, které schválilo vedení úložiště a které popisuje přístup, jímž se má úložiště řídit při uchovávání objektů přijatých do úložiště; pravidla pro uchovávání musí být v souladu se strategickým plánem uchovávání.

**Plán pro zavedení uchovávání** (*preservation implementation plan*) je písemné prohlášení, které schválilo vedení úložiště a které popisuje služby, jež má úložiště nabízet za účelem uchovávání objektů přijatých do úložiště v souladu s pravidly pro uchovávání.

**Procedura** (*procedure*) je písemné prohlášení, které popisuje opatření potřebná k poskytnutí služby nebo k dosažení konkrétního stavu nebo situace; postupy popisují způsob, jakým mají být plněny příslušné plány pro zavedení uchovávání.

---

<sup>12</sup> „Formuluje strategie a postupy dlouhodobé ochrany elektronických dokumentů a provozuje důvěryhodné digitální úložiště.“

Mandát archivu, strategický plán uchovávání a pravidla pro uchovávání lze označit v české organizační terminologii za řídicí dokumentaci; plán pro zavedení uchovávání a proceduru za metodický pokyn.

Norma dále uvádí pojem postup (*practice*), což je opatření sloužící k provádění procedur. Tento postup je hodnocen podle záznamů aplikací nebo jiných dokladů o provedených opatřeních.

### 3 Složky digitalizátů knih v současné praxi

Tato část obsahuje přehled a základní vstupní analýzu současné praxe řízení životního cyklu digitalizátů knih v současných knihovnách, počínaje masovou digitalizací a konče zpřístupňováním čtenářům. Současnou praxí se myslí nejrozšířenější typ digitalizace, který směrnice FADGI definují jako „konverze analogových barevných a jasových hodnot do nespojitých číselných hodnot. Číslo nebo množina čísel označuje barvu a jas každého pixelu v rastrovém obrázku.“ (20, s. 44). Výstupem této digitalizace a předmětem následného uchovávání a zpřístupňování je digitalizát knihy, který je tvořen množinou souborů rastrových dat, jež reprezentují vizuální vlastnosti částí tištěné knihy (stránek) a jejich posloupnost, a obvykle také zahrnuje textová data, která jsou výstupem optického rozpoznávání znaků (OCR) v rastrových obrazech a která umožňují do určité míry pracovat s obrazy jako s textem (jde zejména o funkci plnotextového prohledávání digitalizátu knihy).

Předmětem zkoumání je doba od počátku tohoto milénia, tj. doba, ve které vzešel v obecnou známost model OAIS. Před tímto klíčovým předělem probíhalo řízení životního cyklu digitalizovaných dokumentů bez problematizace otázky digitální archivace. Primárním cílem digitalizace v knihovnách byla zpravidla ochrana tištěných předloh a nová forma zpřístupnění čtenářům (v online dostupných digitálních knihovnách). Součástí plánování ochrany tištěných knih byl koncept digitalizátu knihy jako digitálního surogátu (náhrady původního dokumentu reformátovaným dokumentem) tištěné knihy, který se stane plnohodnotným surogátem v budoucnosti, až dojde ke zničení tištěné předlohy vlivem degradace (urychlené zejména užitím nekvalitního papíru pro vydávání knih v 19. století). V 90. letech 20. století nebyl kladen takový důraz na správný výběr formátu a neexistoval specializovaný metadatový popis, který by zohledňoval požadavky dlouhodobého uchovávání (např. potřebu vytváření interpretačních a provenienčních informací).

V této kapitole je obsažen popis předmětu autenticity (kterým je digitalizát knihy) v současné praxi, jako nutná podmínka pro teoretickou část práce pojednávající o autenticitě tohoto předmětu (digitalizátu knihy). Tato kapitola se zabývá složkami digitalizátu knihy (tj. obsahuje popis předmětu autenticity ze statického hlediska), následující Kapitola 4 pak popisuje etapy životního cyklu digitalizátu počínající jeho vznikem (tj. obsahuje popis předmětu autenticity z dynamického hlediska). Pro přehledovou analýzu v Kapitolách 3 a 4 bylo užito následující pravidlo: primárním zdrojem jsou standardy, digitalizační směrnice a podobné dokumenty popisující osvědčený nebo doporučený postup. Doplněvány jsou

o informace o postupech z konkrétní praxe, a to především v oblastech, které tyto standardy nepopisují (zpravidla jde o velmi specializovanou problematiku).

### 3.1 Formáty

Jedna z definic pojmu formát je „deterministická specifikace toho, jakým způsobem lze vlastnosti nějakého digitálního objektu vratně převést do lineárního bajtového toku (bitového toku)“ (21, s. 15). Americká studie o formátech souborů uvádí: „Většina souborů – s výjimkou souborů, které jsou jednoduchými datovými toky – obsahuje dvě základní komponenty: strukturální prvky a datové prvky. Formát souboru reprezentuje jedinečné a specifické uspořádání těchto strukturálních a datových prvků.“ (22, s. 2)

Formát je jedním z typů interpretačních informací modelu OAIS, konkrétně informací o způsobu, jakým je potřeba datový objekt uložený v konkrétním formátu interpretovat (softwarovou aplikací), a v případě digitalizátů knih také o způsobu, jak datový objekt v daném formátu reprodukovat, tj. adekvátně zobrazit.

V kontextu digitální archivace by v ideálním případě: a) měl vždy být k dispozici dostatek volně dostupných (specializovaných) softwarových aplikací, které znalostí daného formátu disponují (tj. aplikace, které mají dostatečné strukturální interpretační informace) a dokáží s ním pracovat; b) informace o formátu by měly být obsaženy v dostatečně dobře popsané dokumentaci (formátové specifikaci), která by měla být dostupná; c) formát by neměl být zatížen patenty. Počítačová realita má však k tomuto ideálu daleko a právě formáty a aplikace pro práci s nimi jsou předmětem častých změn, a tedy rizikem pro dlouhodobé uchovávání.<sup>13</sup> Jeden z odhadů průměrné délky zastarání formátu (od doby uvedení) na trh se pohybuje v rozmezí osmi až dvaceti let (23). Zastarávání se projevuje dvěma způsoby. Jednak narůstajícím rizikem ztráty dostupnosti softwarových aplikací, které dokáží formáty adekvátně reprodukovat (tj. zobrazit, přehrát nebo jiným způsobem prezentovat smyslům lidského uživatele). Zadruhé pak ztrátou schopnosti nové generace softwarových aplikací formát zpracovávat (upravovat data v daném formátu, provádět konverzi do jiného formátu apod.). V řadě případů také není dostupná formátová specifikace nebo se k formátům váží licenční omezení, což rovněž představuje velká rizika pro uchovávání informací. Připomeňme si druhou povinnost archivu OAIS (získat možnost s informacemi dostatečně nakládat), kterou lze zřejmě aplikovat i na tento případ, neboť nad informacemi uloženými ve formátu,

---

<sup>13</sup> Blíže viz přehled v mé knize (1, s. 50-52)

jehož specifikace je nedostupná (uzavřená) nebo zatížená patenty, z principu nelze získat plnou kontrolu.

Pokud nedojde k včasné formátové konverzi v době, kdy ještě existují vhodné nástroje pro tuto transformaci, může dojít k nevratné ztrátě informačního obsahu (objekt CDO může existovat jako uložený objekt, ale nebude z něj možné získat informační obsah). Archiv by mohl teoreticky vytvořit nový nástroj na základě uložené formátové specifikace zastaralého formátu, prakticky je však takovou alternativu obtížné ověřit.<sup>14</sup>

### 3.1.1 Roviny a aspekty užití formátu

Z hlediska popisu formátu je musíme odlišovat několik rovin, přičemž v případě rastrových formátů jde zejména o tyto roviny:

- Rodina formátů
- Konkrétní formát
- Verze formátu
- Komprese
- Profil

Příkladem rodiny formátů je RAW. Existuje celá řada konkrétních formátů RAW této rodiny, které vytvořili výrobci zařízení. Například fotoaparáty Canon užívají formát Canon RAW. Verze je dána historicky a odlišné verze formátu mohou znamenat odlišné požadavky na zobrazení. Různé formáty nabízejí odlišné možnosti komprese (např. TIFF verze 6 může být v nekomprimované podobě, zatímco JP2 je vždy komprimovaný). Profil znamená specifické nastavení v rámci formátu při jeho vytváření (např. volba ztrátové nebo bezetrátové komprese), u některých formátů nastavení profilu vyžaduje specializovanou znalost (což je zejména případ formátu JP2).

Z hlediska užití musíme u rastrových formátů odlišovat nejméně tyto aspekty: a) archivační formát; b) prezentační formát; c) meziformát.

---

<sup>14</sup> Vzhledem k tomu, že je obtížné předvídat vlastnosti budoucího technologického prostředí, které mohou vytvoření takového nástroje znemožňovat, nebo vzhledem k tomu, že již nebudou dostupné další interpretační informace, které formátová specifikace předpokládala, ale nezaznamenala, neboť v době vytvoření formátové specifikace byly tyto další interpretační informace běžně dostupné.

Archivační formát je takový, který je aktuálně vhodný z hlediska potřeb dlouhodobého uchovávání. Někdy je nazýván jako archivní obrazová matrice (*archival master*). Koncept obrazové matrice (master) byl do digitální archivace digitalizát knih převzat z komerčního sektoru. Obrazová matrice jsou v obrazovém průmyslu obvykle komprimovaná data, která slouží jako zdroj pro vytváření obrazových dat v různé kvalitě, v různých formátech a pro různé účely a nabízí nejvyšší možnou kvalitu dané produkce (např. fotografa). Archivační formát je volen mj. právě s ohledem na to, aby měl tuto funkci obrazové matrice, přičemž však jsou na jeho výběr kladeny další omezující znaky (blíže viz oddíl 3.1.2).

Prezentační formát je takový, který je aktuálně vhodný pro zpřístupňování z hlediska potřeb cílové komunity, v kontextu současné praxe formou prezentace v digitální knihovně. Prezentační formát můžeme rozdělit na hlavní prezentační formát (formát představující nejvyšší možnou kvalitu) a doplňkové prezentační formáty (např. malé náhledy obrázků).

V současné praxi můžeme odlišit čtyři základní typy užití meziformátu, které pro účely této práce nazveme jako konverzní, produkční, prezentační a extrakční meziformát.

Konverzní meziformát je dočasný meziformát užívaný v procesu konverze z nějakého původního formátu do požadovaného cílového formátu. Hlavním příkladem takové konverze ze současné praxe je transformace formátu rodiny RAW do formátu JP2 (formát rodiny RAW musí být nejprve převeden do meziformátu TIFF a teprve z něj je provedena konverze do formátu JP2).<sup>15</sup> Důvodem užití konverzního meziformátu jsou omezení konverzních nástrojů (tj. omezení na vstupní formáty).

Produkční meziformát je dočasný formát, do kterého jsou data v původním formátu převedena za účelem dalšího zpracování (např. provádění obrazových úprav) v etapě digitalizace. Příkladem je převod formátu Canon RAW do formátu DNG. K tomuto převodu zpravidla dochází automaticky (např. při otevření formátu Canon RAW v aplikacích Adobe dojde k jeho automatickému převodu do formátu DNG). V současné praxi se formát DNG následně převádí do formátu TIFF jako archivačního formátu.<sup>16</sup>

Prezentační meziformát je meziformát, ze kterého digitální knihovna generuje cílový prezentační formát. Nejčastějším případem současné praxe je formát JP2 ve ztrátové kompresi, ze kterého se v digitální knihovně generuje formát JPEG jako výsledný hlavní prezentační formát. Jednou z výhod JP2 jako prezentačního meziformátu je řešení problému,

---

<sup>15</sup> Tento postup užívá například Wellcome Library (117).

<sup>16</sup> Tento postup užívá například švédská národní knihovna (34).

že formát JP2 není nativně podporován v běžných internetových prohlížečích, a rovněž relativní jednoduchost takovéto formátové konverze do JP2.

Extrakční meziformát je dočasný meziformát, který je určen jako zdroj pro proces OCR. Důvodem vytvoření tohoto meziformátu může být zjištění, že proces OCR přináší mírně lepší výsledky při užití komprimovaném formátu jako zdroje (24, s. 3).

### 3.1.2 Výběr archivačního formátu

Široce známým doporučeným postupem pro řízení životního cyklu digitálních dokumentů je, že prvním a zásadním krokem je výběr vhodného archivačního formátu, tj. formátu, ve kterém budou dokumenty uchovávány v archivu (a v případě, kdy dokumenty vytváří i uchovává jedna a tatáž organizace, vytváření finálních dat přímo v tomto formátu). Archivační formát je volen z hlediska jeho (aktuální) vhodnosti pro uložení v balíčku AIP v digitálním archivu. Cílem je uložit obsah v takovém formátu, o kterém se předpokládá, že jeho užití v současnosti a blízké budoucnosti nebude představovat větší riziko. Pokud vkladatel do archivu dodá data v jiném formátu než archivačním, je doporučeným postupem, aby archiv provedl normalizaci do archivačního formátu (1, s. 83-86).

Volba archivačního formátu není jednoduchá záležitost, nicméně dnes již existuje řada doporučení od uznávaných organizací, kterými se lze řídit. Formátový registr Kongresové knihovny již dlouho patří mezi hlavní zdroje pro stanovení kritérií výběrů formátů (25). Z hlediska seznamu doporučených konkrétních formátů byl ještě před několika lety de facto jediným citovaným zdrojem v odborné literatuře dokument Floridského digitálního archivu (Florida Digital Archive). Tento dokument obsahoval výčet nejběžnějších formátů s hodnocením jejich spolehlivosti na třístupňové škále (vysoká, střední a nízká spolehlivost) pro potřeby uchovávání v tomto archivu.<sup>17</sup> Od té doby vzniklo více takových doporučení, která lze považovat za směrodatná, neboť je vydaly významné organizace.

Za jeden z nejvýznamnějších zdrojů pro výběr archivačního formátu lze v současnosti označit dokument Deklarace doporučených formátů (Recommended Formats Statement), který od roku 2014 vydává Kongresová knihovna (Library of Congress). Tento dokument obsahuje konkrétní seznam doporučených formátů (pro digitální i fyzické objekty), který je určen jak pro vnitřní potřeby Kongresové knihovny, tak i pro jiné archivy. Zahrnuje

---

<sup>17</sup> Dokument nesl název „Recommended Data Formats for Preservation Purposes in the Florida Digital Archive“ a v současnosti již není dostupný (1, s. 84).



seznamy preferovaných a akceptovaných formátů. Dokument je každý rok aktualizován, poslední seznam byl vydán v červenci 2016 (26).

Dalším významným doporučením je studie americké iniciativy FADGI, která byla vydána v roce 2014 a zaměřuje se specificky na rastrové formáty (tedy formáty relevantní pro digitalizaci knih) (27). Studie obsahuje zhodnocení nejběžnějších formátů (TIFF, JP2, PDF, PNG, JPEG) z hlediska digitální archivace a uvádí sadu podrobných srovnávacích kritérií seskupených do čtyř hlavních kategorií (udržitelnost, ekonomické faktory, požadavky na implementaci, nastavení a možnosti).

Kritéria, která se v různých doporučeních objevují nejčastěji, lze zobecnit do těchto kategorií: podpora, otevřenost a nezatíženost patenty.<sup>18</sup> Podporou formátu se zde rozumí míra jeho užívanosti v dané komunitě a dostupnost nástrojů pro vytváření, zpracování a reprodukci. Nízká úroveň podpory znamená, že formát zastarává nebo že se vůbec neujal. Otevřeností formátu se rozumí skutečnost, že je dostupná dokumentace formátu (tj. formátová specifikace). Směrnice MINERVA uvádějí, že užití otevřených formátů „napomůže interoperabilitě a zajistí, že zdroje lze opětovně využívat, vytvářet a upravovat celou řadou aplikací. Rovněž napomůže vyhnout se závislosti na konkrétním dodavateli.“ (28, s. 32)

Za dostupnost se v praxi zpravidla považuje to, že formátová specifikace je buď volně dostupná online, nebo že je dostupná v nějaké normalizační organizaci, která ji udržuje. Například specifikace formátu TIFF je dostupná na webu jeho vlastníka (firmy Adobe) a je průmyslovým standardem. Specifikace formátu JP2 je mezinárodní normou, která je dostupná k zakoupení v normalizační organizaci ISO (29).

Nezatíženost patenty nutně neznamená, že formát není nikým vlastněn, pouze to, že výkon práv duševního vlastnictví není uplatňován.

Dalšími uváděnými kritérii pro výběr formátu jsou například: míra zpětné kompatibility; možnosti exportu do jiných formátů; míra nezávislosti na specifických hardwarových a softwarových platformách; rozumná rovnováha mezi nabídkou funkcí formátu na straně jedné a přiměřenou komplexitou na straně druhé (1, s. 85).

Obecně lze říci, že archivační formát by měl splňovat mj. také funkce obrazové matrice. Mimo oblast digitální archivace je častou volbou (zejména profesionálních fotografů) uložení obrazové matrice ve formátu DNG. Takovéto užití je v digitální archivaci problematické jednak kvůli svázanosti formátu DNG s aplikacemi firmy Adobe (která

---

<sup>18</sup> Srv. např. (1, s. 83), (28, s. 11-14), (115).

je původcem tohoto formátu), jednak také k nemalé ceně těchto aplikací, která může být pro řadu paměťových organizací zásadní provozní překážkou. To ukazuje, jak již bylo uvedeno výše, že požadavky na archivační formát jdou nad rámec požadavků na obrazovou matici.

### 3.1.3 Doporučované a užívané archivační formáty

Směrodatné zdroje za dva hlavní archivační formáty uvádějí TIFF a JP2. Tyto formáty splňují uvedené tři základní podmínky (podpora, otevřenost, nezátíženost patenty). Je jistou výhodou, že v oblasti rastrových formátů je možnost takové volby vzhledem k tomu, že existují typy dat, pro které otevřené formáty zatím neexistují nebo se neužívají.

Směrnice FADGI doporučují formát TIFF, verze 6, profil bez komprese (20, s. 79). Doporučení Floridského digitálního archivu označovalo za rastrové formáty s nejvyšší spolehlivostí nekomprimovaný TIFF, JP2 v bezztrátové kompresi a PNG.<sup>19</sup> V Deklaraci doporučených formátů Kongresové knihovny pro 2016-2017 se pro digitální obrazová data (digitální fotografie a další typy digitálních obrazových dat) uvádějí jako preferované archivační formáty: 1. TIFF, 2. JP2, 3. PNG (26). Směrnice pro budování kvalitních digitálních sbírek (Framework of Guidance for Building Good Digital Collections) vydané americkou normalizační organizací NISO doporučují jako archivační formáty nekomprimovaný TIFF a bezztrátově komprimovaný JPEG 2000 (30, s. 28). Podle rozsáhlého průzkumu provedeného v letech 2012-2013 považují severoamerické knihovny formát TIFF za nejspolehlivější archivační formát vůbec (bez ohledu na typ dokumentu) (31).

V současné praxi je nejužívanějším archivačním formátem TIFF.<sup>20</sup> Nekomprimovaný TIFF jako archivační formát využívají pro své digitalizační projekty například Kongresová knihovna (32, s. 5), nebo národní knihovny Francie (33) a Švédska (34). Převaha formátu TIFF jako archivačního formátu digitalizačních projektů souvisí nejen s důvěrou, které se těšil a stále těší, ale také s jeho dlouhou historií (specifikace současné aktuální, tj. šesté verze byla vydána v roce 1992) a širokou podporou v technologickém prostředí od 90. let 20. století po současnost.

Přibližně posledních deset let se začíná v novějších digitalizačních projektech zejména evropské provenience rozšiřovat užití formátu JP2 jakožto archivačního formátu (35), (32). Jistý vliv na to může mít i příznivá studie italských odborníků, která byla v roce 2008 vydána v periodiku D-Lib Magazine (36). Tato studie srovnávala tři rastrové formáty (TIFF, JP2

---

<sup>19</sup> Citováno dle (1, s. 84).

<sup>20</sup> Srv. např. (32, s. 3-4), (31), (35)

a JPEG) z hlediska jejich vhodnosti pro archivaci. Za nejvhodnější archivační formát označila právě JP2, mj. díky jeho nejlepší robustnosti (odolnosti vůči menšímu poškození bitů). Formát JP2, který byl původně v knihovnách užíván spíše jako formát pro zpřístupnění, se stal archivačním formátem například pro digitální obrazové fondy knihovny Wellcome Library (37) nebo pro masovou digitalizaci norské národní knihovny (38) či projekt NDK Národní knihovny ČR (39). Poslední dvě jmenované knihovny užívají profil formátu JP2 v bezztrátové kompresi, zatímco první používá ztrátovou kompresi. Užití ztrátové komprese není výjimečné. V praxi jsou voleny různé profily formátu JP2, zahrnující nejen volbu typu komprese, ale některé další parametry specifické pro tento formát (40). Vhodné nastavení těchto parametrů vyžaduje specialistu; z toho důvodu si některé knihovny nechaly vypracovat profily na zakázku u externích odborníků.<sup>21</sup> V posledních letech bylo uspořádáno několik seminářů věnujících se této problematice. K formátu JP2 se však objevují i kritické názory (35) a diskuze o tomto formátu jako archivačním formátu pokračuje (32, s. 3). Obecně lze říci, že v USA převažuje volba formátu TIFF a v Evropě se rozšiřuje užití formátu JP2. Užití formátu PNG jako archivačního formátu v současné praxi je řídké, což může být vzhledem k jeho kvalitám překvapivé.

#### **3.1.4 Prezentační formáty**

Volba prezentačního formátu závisí na požadavcích cílové komunity. Jelikož v současné praxi je hlavní formou zpřístupnění digitalizátů jejich zobrazení v digitální knihovně, tyto požadavky se řídí především podporou formátu v internetových prohlížečích (41, s. 46). Směrnice MINERVA doporučují zpřístupňovat digitální kopie dokumentů v různých velikostech nebo formátech, aby zacílení na uživatele bylo co nejširší (28, s. 73). Směrnice pro budování kvalitních digitálních sbírek (Framework of Guidance for Building Good Digital Collections) vydané americkou normalizační organizací NISO doporučují pro zpřístupňování digitalizovaných dokumentů užití formátů JPEG a PDF.

V praxi je nejužívanějším hlavním prezentačním formátem digitálních knihoven JPEG (32). Tato volba je logická vzhledem k jeho vysoké podpoře v internetových prohlížečích nebo zobrazovacích aplikacích. Formát JPEG jako prezentační užívá například Gallica, jedna z největších digitálních knihoven světa provozovaná francouzskou národní knihovnou (zde je tento prezentační formát vytvářený konverzí z formátu TIFF, který je archivačním formátem této knihovny) (33, s. 5). Formát JP2 není podporován běžnými internetovými prohlížeči,

---

<sup>21</sup> Například profil formátu JP2 pro Wellcome Library vytvořil Robert Buckley, jeden z autorů specifikace formátu JP2 (37).

pro jeho užití jako prezentačního formátu je nutno nainstalovat plugin. Běžným způsobem zpřístupnění, který řeší tento problém z hlediska komfortu cílové komunity, je využití formátu JP2 jako prezentačního meziformátu, ze kterého digitální knihovna generuje formát JPEG, v němž jsou obrazová data prezentovaná čtenářům internetovým prohlížečem (37, s. 11). Jako doplňkové vedlejší formáty se užívají například formát PDF (pro možnost stažení digitalizátu knihy čtenářem v podobě jednoho souboru) nebo formát GIF (pro náhledy obrázků).

### 3.1.5 Formátové registry

V souvislosti s přijetím modelu OAIS se v komunitě paměťových institucí objevil navazující koncept globálního registru interpretačních informací.<sup>22</sup> Idea takového registru spočívá v tom, že bude sloužit jako zásobník interpretačních informací, které jsou potřebné pro digitální objekty (zejména informací o formátech, souvisejících aplikacích a všech ostatních prvcích počítačového prostředí, jež podporuje adekvátní reprodukci digitálních objektů a jejich zpracování). Za tímto návrhem stála pragmatická úvaha, podle níž nejsou jednotlivé instituce schopny všechny potřebné interpretační informace shromažďovat vlastními silami.

Jako první vznikl registr PRONOM (42). Byl založen a je již patnáct let provozován britskými Národními archivy (The National Archives). Ačkoliv PRONOM zdaleka nesplňuje své původní ambice (obsahuje poměrně rozsáhlou databázi formátů, ale jejich popis je většinou minimální), jde v současnosti o nejvýznamnější důvěryhodný projekt, který se alespoň snaží uskutečňovat původní vizi směrodatného globálního registru interpretačních informací. Později sice vznikly dva další registry, ale již zanikly. Prvním z nich byl GDFR, který skončil již před několika lety.<sup>23</sup> Registr UDFR provozovaný Kalifornskou digitální knihovnou (California Digital Library) měl spojit registry GDFR a PRONOM, ale byl ukončen v dubnu 2016 kvůli nedostatku financí.<sup>24</sup>

Klíčovou funkcí registru PRONOM je to, že (jako jediný registr vůbec) nabízí jednoznačný a jedinečný identifikátor formátu, přesněji řečeno, jak uvádí sám registr: „rozšiřitelné schéma pro poskytování perzistentních, jedinečných a jednoznačných identifikátorů pro jednotky interpretačních informací zaznamenané v registru PRONOM.“ (43, s. 4). Formát je tedy pouze jedním z typů interpretačních informací, o nichž registr vede

---

<sup>22</sup> Srv. např. (112).

<sup>23</sup> [http://library.harvard.edu/preservation/digital-preservation\\_gdfr.html](http://library.harvard.edu/preservation/digital-preservation_gdfr.html)

<sup>24</sup> <http://www.udfr.org/>

údaje, nicméně nejrozšířenějším. Funkce identifikátoru PUID jsou dvě: propojení se záznamem jednotky interpretačních informací v registru PRONOM (tj. způsob identifikace záznamu, přičemž tento záznam by ideálně měl obsahovat co nepodrobnější informace o formátu nebo jiné jednotce interpretačních informací) a jedinečný perzistentní identifikátor, který odlišuje v maximální možné míře jeden formát od druhého (odlišuje se nejen typ formátu, ale i verze). Například PUID pro JPEG verze 1.00 je „fmt/42“, pro verzi 1.01 „fmt/43“ a pro verzi 1.02 „fmt/44“. Registr MIME,<sup>25</sup> který je nejužívanějším obecným registrem formátů (sloužícím i pro účely mimo kontext digitální archivace), odlišuje formáty jen na základě typu a názvu, například formát JPEG všech verzí má označení „image/jpeg“.

Za druhý významný zdroj informací o formátech lze považovat registr Kongresové knihovny (25). Ten obsahuje nejen základní interpretační informace o formátech, ale také o rizicích s nimi spojených.

Formátový registr PRONOM nabízí záznamy o formátech a dalších jednotkách interpretačních informací v podobě volně dostupných webových stránek. Současně poskytuje identifikační mechanismus, který obsahuje popis toho, kde se v daném formátu nachází tzv. „magické číslo“ (údaj o verzi formátu), a identifikátory PUID. „Magické číslo“ je interní mechanismus označení konkrétního formátu daný formátovou specifikací, zatímco identifikátor PUID je jedinečný externí identifikátor, který je přidělován registrem PRONOM. Registr PRONOM také nabízí svůj vlastní nástroj, DROID, který provádí formátovou identifikaci užitím uvedeného mechanismus na jednotlivé soubory.

## 3.2 Obrazová data

Obrazová data jsou klíčovou složkou digitalizátů knih současné praxe. Jak již bylo uvedeno, v současné praxi jsou tištěné knihy snímány (skenovány nebo fotografovány) do podoby rastrových obrazových dat. U obrazových dat je dále třeba rozlišovat několik aspektů nebo vlastností, mezi které patří rastrový formát a dále: generace dat; strukturální model obrazové reprezentace; formátový profil; typy komprese; obecné obrazové vlastnosti; prezentační varianty obrazových dat.

---

<sup>25</sup> <http://www.iana.org/assignments/media-types/media-types.xhtml>

### 3.2.1 Generace obrazových dat

Prvotní výstup snímání skenerem nebo fotoaparátem budeme nazývat původní snímky. Původní snímky jsou soubory v rastrových formátech uložené po snímání na datový nosič pracovní stanice digitalizační linky. Tyto původní snímky zpravidla procházejí dalším zpracováním do doby, než je vytvořen konečný obrazový výstup digitalizace (finální produkční data). Každá transformace, kterou prošla obrazová data od původních snímků po vytvoření finálních produkčních dat, znamená vytvoření nové generace obrazových dat. K vytváření dalších generací dochází také v pozdějších etapách (archivace a zpřístupňování), nicméně cíle vytváření těchto generací jsou odlišné. Cílem digitalizačních transformací je vytvořit konečný digitální produkt a teprve tento konečný produkt lze považovat za plnohodnotnou obrazovou složku digitalizátu knihy. Cílem archivace a zpřístupnění je zachovat tento digitalizační produkt v požadované kvalitě, v případě archivace z hlediska uchovávání informačního obsahu navzdory rizikům technologického zastarávání, v případě zpřístupnění vytvoření takové podoby digitalizátu dat, která je vhodná pro aktuální potřeby cílové komunity a která se může lišit od dat uložených v balíčku AIP. Původní snímky v současné praxi mohou být v odlišném formátu než finální produkční data. Generace se od sebe liší změnami, které lze zaznamenat na bitové úrovni, přičemž však může platit, že některé transformace je možno vykonat společně před tím, než bude uložena nová generace obrazových dat.

### 3.2.2 Strukturální model obrazové reprezentace

Modelem obrazové reprezentace se zde rozumí vztah mezi strukturou tištěné knihy (posloupností stran) a způsobem digitální reprezentace této předlohy v digitalizátu knihy. Převažující strukturální model současné praxe lze charakterizovat následovně:

- Základní předmět reformátování = jedna stránka knihy.
- Základní objekt uložení = jeden soubor reprezentující tuto jednu stránku.

Tento model znamená, že jeden soubor v rastrovém formátu reprezentuje jednu stránku knižního bloku, přičemž všechny rastrové soubory mají stejnou velikost. Pro reprezentaci jiných částí knih, než jsou stránky (např. přebal), nebo nestandardních částí (např. stránka s mapou, kterou lze rozložit, takže její velikost bude jiná než ostatních stran), se v praxi užívají různé postupy, kterými se v této práci nebudeme z důvodu zjednodušení zabývat. Běžnou současnou praxí je snaha zachytit všechny části knihy, včetně vakátů

a prázdných přídeští (jedinou výjimkou v tomto směru je zadní část přebalu, která bývá bílá a nepovažuje se za smysluplné ji digitalizovat).

### **3.2.3 Formátový profil**

Formátovým profilem se rozumí nastavení v rámci konkrétního formátu. Mezi hlavní prvky profilu rastrových formátů patří komprese (tj. její nastavení v rámci možností formátu, přičemž ne všechny rastrové formáty umožňují všechny typy kompresí (viz oddíl 3.2.4). Nejčastějším profilem pro formát TIFF je nekomprimovaná varianta. Formátový profil pro JP2 zahrnuje (nutnou) volbu mezi ztrátovou a bezztrátovou kompresí a dále několik dalších specifických nastavení typických pro tento formát.

### **3.2.4 Typy komprese**

Klíčovým aspektem rastrových obrazových dat je komprese. Pro potřeby této práce postačí následující klasifikace čtyř základních typů:

1. Nekomprimovaná varianta (tj. komprese není užitá)
2. Matematicky bezztrátová komprese
3. Vizuálně bezztrátová komprese
4. Vizuálně ztrátová komprese

Nekomprimovaná varianta může být teoreticky nejlepší možnou volbou pro archivační formát. Studie iniciativy FADGI uvádí: „Nekomprimovaná datová struktura má jednu velkou výhodu: je relativně transparentní. Transparentnost souvisí s ukazatelem udržitelnosti: nemělo by být složité vytvořit nástroj, který dokáže přečíst informaci o obalu (wrapper) a rozbalit rastrová data tak, aby je bylo možno zobrazit“ (32, s. 4). Jediným problémem s užitím nekomprimovaných dat je jejich velikost. V současné praxi je však tento problém někdy velkou překážkou vzhledem k omezenému rozpočtu na úložné kapacity.

Matematicky bezztrátová komprese obrazových dat v principu odpovídá konceptu vratné transformace modelu OAIS. Kompresní algoritmus snižuje transparentnost formátu, ale současně umožňuje snížení požadavků na úložné kapacity.

Vizuálně bezztrátovou kompresí se myslí taková matematicky ztrátová komprese, která na základě nějakého přijatého psychofyziologického modelu stanoví kompresní poměr, jehož výstupem (při zobrazení) má být informační obsah, který by lidský vnímatel neměl rozeznat od výstupu matematicky bezztrátové komprese, nebo jsou viděné rozdíly nepodstatné (37, s. 4).

Vizuálně ztrátová komprese je taková matematicky ztrátová komprese, která přináší vizuálně patrné změny obrazové kvality. Míra komprese v rámci tohoto typu se může dále výrazně lišit, přičemž nejvyšší možnou vizuálně ztrátovou kompresi lze pochopitelně užít pouze pro náhledy obrázku (tedy pro doplňkové prezentační formáty).

Ne všechny rastrové formáty umožňují všechny výše uvedené volby. Formát TIFF umožňuje všechny varianty: nekomprimovanou variantu (která je také nejčastěji užívaná pro TIFF jako archivační formát), matematicky bezztrátovou kompresi (algoritmus LZW nebo ZIP) a ztrátovou kompresi (algoritmus JPEG). Formát JPEG nabízí pouze matematicky ztrátovou kompresi (typy 3-4).<sup>26</sup> Formát PNG nabízí pouze bezztrátovou kompresi, podle studie FADGI s vynikajícími výsledky (32, s. 3). Formát JP2 nabízí pouze komprimované varianty (typy 2-4). Koncept vizuálně bezztrátové komprese (typ 3) je v současné praxi spojován právě s tímto formátem. Skupina výzkumníků z několika významných knihoven světa provedla rozsáhlý test vnímání (různě vysoké) ztrátové komprese formátu JP2 mezi čtenáři s cílem navrhnout vhodné profily formátu JP2 ve vizuálně bezztrátové kompresi (24). Některé knihovny pak začaly v praxi využívat vizuálně bezztrátovou kompresi pro archivační formáty.<sup>27</sup> Pro matematicky bezztrátovou kompresi se udává, že kompresní poměr je obvykle zhruba 2:1 (44, s. 6).

### 3.2.5 Obecné obrazové vlastnosti

Směrnice FADGI určují jak čtyři základní obrazové vlastnosti rastrových dat prostorové rozlišení (*spatial resolution*), bitovou hloubku (*bit-depth*), barevný model (*color model*) a barevný profil (*color profile*).

Prostorové rozlišení určuje množství informací v rastrovém souboru z hlediska počtu pixelů (obrazových prvků) na jednotku měření, obvykle palec (odtud zkratka PPI),<sup>28</sup> tj. „stanovuje, jak blízko od sebe jsou jednotlivé pixely umístěny“; bitová hloubka „stanovuje maximální počet odstínů nebo barev v digitálním obrazovém souboru“ (20, s. 4). Prostorové rozlišení a rozměry digitálního obrazu určují celkový počet pixelů v souboru; při určení požadované velikosti souboru je nutno zadat hodnotu prostorového rozlišení a rozměry (např. 300 ppi + 8x10 palců). Současným standardem pro rastrové obrazy ve stupních šedi i barevné podobě je užití bitové hloubky 8 bitů pro jeden pixel pro jeden kanál. Obrazy ve stupních šedi

---

<sup>26</sup> Srv. (36).

<sup>27</sup> Například Wellcome Library (37)

<sup>28</sup> Zkratka pro počet pixelů na palec (pixel per inch), někdy se užívá též počet pixelů na milimetr nebo centimetr.



mají jeden kanál, barevné obrazy tři a více kanálů. Barevný model je způsob číselné specifikace barev s užitím tří nebo více kanálů. Například barevný model RGB obsahuje tři kanály o bitové hloubce 8 nebo 16 bitů.

Prostorové rozlišení tedy vymezuje, jak detailně může být převedena tištěná kniha (např. čitelnost písma), a bitová hloubka, jak věrně mohou být zachyceny její barvy nebo odstíny šedi (tedy barevná věrnost) ve výsledném souboru. Prostorové rozlišení, ani bitová hloubka logicky nemohou udávat ani zaručovat kvalitu uložených informací, pouze vymezují rozsah možné kvality digitalizačního převodu. Kvalita digitalizátu se odvíjí od míry detailnosti (např. velikost písma) a barevnosti předlohy, která je digitálně zachycována, a nastavených hodnot prostorového rozlišení a bitové hloubky. Obecně platí, že rozmezí prostorového rozlišení a bitové hloubky je na jedné straně ovlivněno minimální hranicí (tj. aby kniha ještě byla čitelná a její obrazové prvky do určité míry rozpoznatelné), na druhé straně maximální hranicí rozlišení, tj. takové, nad jejíž rámec snímání nemůže v principu přinést již žádný pozorovatelný rozdíl, a digitalizace ve vyšším rozlišení by byla neekonomická či jinak neúčelná. Existující doporučení se tedy mohou pohybovat pouze mezi těmito krajními případy. Doporučení současné praxe však také zohledňují skutečnost, že při masové digitalizaci je prakticky nemožné vytvářet tato nastavení pro každou knihu zvlášť. Z tohoto důvodu se vydávají plošná doporučení pro minimální rozlišení a bitovou hloubku pro různě definované kategorie knih z hlediska jejich předpokládané velikosti písma a barevnosti.

Pro běžné knihy novodobých fondů obsahující barevné prvky se obecně doporučuje barevný model RGB, bitová hloubka 24 bitů a minimální prostorové rozlišení 300-400 PPI (20) (45). Digitalizační projekt může vytvořit odlišné pracovní postupy například pro tištěné knihy obsahující barevné ilustrace nebo fotografie a tištěné knihy obsahující pouze text (a druhé snímat jen ve stupních šedi). Při masové digitalizaci však může být z hlediska přípravy obtížné kontrolovat, zda kniha neobsahuje barevné prvky. Z tohoto důvodu se často vytváří jedno plošné nastavení obrazových vlastností pro všechny knihy.<sup>29</sup>

V případě prostorového rozlišení nemohou být doporučení minimálního rozlišení ničím jiným než predikcí očekávané nejmenší velikosti písma u určité skupiny předloh. V případě, že se ve skupině vyskytne anomálie, pak tato plošná doporučení přirozeně nemohou zaručit kvalitní výsledek snímání. Hodnota prostorového rozlišení se nastavuje

---

<sup>29</sup> Příkladem je digitalizace projektu NDK, kde je plošně pro všechny digitalizované dokumenty nastaveno rozlišení 300 PPI, barevný model RGB, bitová hloubka 8 bitů (39).

pouze na skeneru, pro fotografování tento údaj přirozeně nemá smysl. Na výsledné rozlišení fotografie má vliv kvalita senzoru fotoaparátu a objektivu, vzdálenost knihy od objektivu a zaostření.

Zatímco nastavení výše uvedených vlastností (prostorové rozlišení, bitová hloubka a barevný model) v současné praxi nepředstavuje větší problém, barevný profil je složitější problematika. Barevný profil „určuje interpretaci číselných hodnot popisujících pixely v obrázku tím, že popisuje chování zařízení nebo rozsah barevného prostoru“ (20, s. 45). Barevný prostor je „geometrická reprezentace barev v prostoru, který lze vizuálně vnímat nebo vytvářet užitím konkrétního barevného modelu“ (46). Barevný prostor je například vyžadován aplikacemi pro zobrazení. Snímací zařízení zpravidla užívají vlastní barevný profil, který je závislý na konkrétním zařízení nebo výrobcí. Tento technologicky závislý profil lze převést do ICC profilu (standardu pro univerzální barevnou specifikaci) a uložit do obrázku.

### **3.2.6 Prezentační varianty**

Hlavním cílem současné praxe je vytvoření finálních produkčních dat v archivačním formátu. Obrazová data v archivačním formátu se následně v nezměněné podobě uchovávají v archivu do doby, než bude z důvodů zastarávání technologií nutno přistoupit k archivačním opatřením (blíže viz oddíl 4.2). Běžnou praxí je, že se v digitální knihovně čtenářům nezpřístupňují obrazová data v archivačním formátu, ale vytváří se jejich prezentační varianta v prezentačním formátu. V případě užití archivačního formátu TIFF bývají prezentační variantou zpravidla obrazová data ve formátu JPEG jako hlavním prezentačním formátu,<sup>30</sup> v případě archivačního formátu JP2 bývá vytvořen prezentační meziformát ve formátu JP2 (ve ztrátové kompresi), ze kterého systém digitální knihovny vytváří za chodu prezentační varianty ve formátu JPEG.

Důvody vytváření prezentační varianty pro zpřístupnění jsou různé a zpravidla jsou kombinací více faktorů. Zaprvé jsou dány historicky. Digitalizace se prováděla již v dobách, kdy bylo internetové připojení ještě pomalé nebo nákladné, a tudíž prezentační varianta tvořená obrazovými soubory menší velikosti (zejména ve ztrátové kompresi) byla uživatelsky vhodnou formou prezentace. Druhým důvodem je, že formát TIFF, hlavní volba digitalizačních projektů, nebyl a dosud není podporován internetovými prohlížeči. Třetím důvodem může být následování praxe obrazového průmyslu, kdy je obvyklým způsobem

---

<sup>30</sup> Srv. např. (66, s. 10), (67, s. 64)

ukládat obrazové matrice v nejvyšší možné kvalitě a užívat je jako zdroj pro generování obrazových dat v různé kvalitě pro různé účely (mj. také prezentace na webu, např. v online periodicích). Odůvodněním také mohou být výzkumy stanovující psychofyzilogický model, podle něhož je informační obsah reprodukováný z obrazových dat ve ztrátové kompresi určité úrovně čtenářem vizuálně nerozeznatelný od informačního obsahu reprodukováného z obrazových dat v bezeztrátové kompresi nebo nekomprimované podobě.<sup>31</sup> Z toho se vyvozuje, že zpřístupňování obrazových matric v archivním formátu je neúčelné. Nikoliv výjimečným případem současné praxe je, že se prezentační varianty vytvářejí již při produkci, jako součást finálních produkčních dat (a tedy nikoliv až v archivu), a to z důvodu jednoduššího zpracování.<sup>32</sup>

### 3.3 Textová data jako výstup OCR

Optické rozpoznávání znaků (OCR) je metoda získávání textu z obrazu. V současné praxi se užívá k tomu, aby se z rastrových dat vzniklých digitalizací vytěžil textový obsah. Výstup z OCR se ukládá v podobě strukturovaného textového formátu, který obsahuje informace o pozici (obrazem vyjádřených) konkrétních písmen (slov) v obrazovém souboru, z něhož byl vytvořen, aby bylo zajištěno namapování textu na obraz. Tímto formátem je v současné praxi převážně ALTO. Formát ALTO sám sebe popisuje jako „standardizovaný formát XML k ukládání informací o rozložení (layout) a obsahu“ (47). Formát ALTO XML je navržen jako externí schéma pro standard METS (o standardu METS blíže v oddíle 3.4). Jde však především o datový formát (obsahuje vlastní text a jeho strukturu), přičemž malá část jeho obsahu jsou metadata. V případě, že se vytvářejí archivační formát i prezentační varianty, je nutné, aby měly stejnou pixelovou velikost (počet pixelů na šířku a výšku obrázku, jinak text nebude sedět).

Míra efektivity OCR závisí na několika faktorech: stav předlohy, kvalita softwaru a jeho slovníků a komprese. Podle rozsáhlého výzkumu přinášejí ztrátově komprimovaná obrazová data mírně lepší výsledky procesu OCR v porovnání s bezeztrátovou kompresí. V praxi platí, že pro novodobé fondy v dobrém stavu předloh (a pro běžné jazyky) je efektivita OCR velmi vysoká. Například podmínkou pro digitalizační projekty financované českým národním dotačním titulem VISK 7 je úspěšnost OCR stanovena na minimálně 95% (48).

---

<sup>31</sup> Viz např. (24)

<sup>32</sup> Viz např. (39).

Prezentace digitalizátu knihy v digitální knihovně zahrnuje nejen zobrazení, ale i reprodukci těchto strukturovaných textových informací získaných z OCR, která umožňuje čtenářům plnotextové prohledávání obrazových dat. Přidáním této strukturované textové složky je digitalizát knihy obohacen o funkci, kterou jeho tištěná předloha nikdy neměla.

### 3.4 Metadata

Norma ISO 14721 definuje metadata jako „data o jiných datech“ (10, s. 24). Tato klasická definice je sice velmi obecná, ale nutná pro to, aby bylo možno jako metadata označit celou škálou odlišných užití a typů metadat, mezi které patří klasické bibliografické záznamy, strukturovaná metadata uložená v samostatných souborech nebo metadata ukládaná do samotných obrazových souborů, která při běžném zobrazení nejsou viditelná.

V současné praxi je nejdůležitější druhý uvedený typ metadat. Tato metadata jsou vysoce strukturovaná, uložena do samostatných textových souborů ve formátu XML, uchovávaná spolu s obrazovými daty a pečlivě zdokumentovaná mezinárodními standardy, které jsou široce užívány celou knihovnickou komunitou. Dodejme, že tento typ metadat má svůj původ v katalogizačních standardech, původně vytvářených pro popis tištěných knih a dalších typů fyzických dokumentů za účelem možnosti jejich strojového zpracování.

Klíčovým prvkem těchto standardů je sada elementů (*element set*). Jak uvádí Zeng, metadatové standardy založené na sadě elementů obecně slouží k popisu zdrojů specifického typu nebo pro konkrétní účel, definují význam elementů a jejich vztahy a poskytují návod, jaké hodnoty a jakým způsobem by měly být připsány elementům při popisu konkrétního zdroje nebo účelu. Každý element je definován určitým počtem atributů (základním je název) a dalších nezbytných informací (definice, identifikátor<sup>33</sup> atd.). Z tohoto důvodu je klíčové, aby všechny sady elementů byly definovány ve formátu, který je konzistentní, srozumitelný a komunikovatelný mezi různými komunitami (49, s. 38). Užití serializace digitálních metadat do formátu XML je podle Zenga typické pro paměťové instituce obecně (49, s. 131).

Pro ilustraci aktuálně vnímané důležitosti úlohy metadat (výše uvedeného typu) v současné praxi poslouží následující citace, která pocházející přímo z jednoho z takových standardů: „Bez strukturálních metadat jsou obrázek stránky nebo textový soubor tvořící

---

<sup>33</sup> Například PREMIS má u každého elementu uveden jedinečný identifikátor založený na hierarchické řazení, např. 1.2, 3.1.

digitální dílo téměř k ničemu, a bez technických metadat zohledňujících digitalizační proces si badatelé nemohou být jisti, jak přesný odraz originálu digitální verze poskytuje. Pro účely vnitřní správy musí mít knihovna přístup k náležitým technickým metadatům, aby mohla pravidelně obnovovat a migrovat data, a tak zajistit zachování cenných zdrojů.“ (50)

### **3.4.1 Užívané metadatové standardy**

Již před érou digitálních metadat byla knihovnická komunita známá vysokou mírou jednotné standardizace na mezinárodní úrovni. Tak tomu je i v případě metadatových standardů pro digitální data. Knihovny pro správu digitalizátů se nejčastěji užívají tyto mezinárodní metadatové standardy: METS (51), PREMIS (52), MODS (53), MIX (54) a ALTO (47) .

Tyto standardy lze z hlediska rozsahu jejich možné aplikace rozdělit do dvou skupin: a) omezené užití (MIX; ALTO), univerzální užití (METS, PREMIS, MODS). První skupina je určena specificky pro digitalizáty knih (a některé další typy digitálních dokumentů), druhou skupinu lze aplikovat na většinu typů digitálních dokumentů. Všechny tyto standardy, s výjimkou PREMIS, jsou definovány specificky jako XML schémata. Standard PREMIS není svázán se schématem XML, ale lze jej jako XML vyjádřit (existuje oficiální XML schéma k tomuto standardu) a jako XML schéma se také užívá pro balíčky SIP a AIP.

Standard ALTO je primárně datový formát (zaznamenávající text získaný procesem OCR a jeho souřadnicové umístění vzhledem k obrazu), ale obsahuje některé metadatové prvky (např. informace o obrazovém zdroji pro OCR). Standard MODS je užíván pro zápis bibliografických informací, a je tedy určen k naplňování role identifikačních informací OAIS (a případně popisných informací OAIS). Zbylé tři standardy, které jsou nejdůležitější, popíšeme detailněji v následujících oddílech.

### **3.4.2 Metadatový aplikační profil**

Každá organizace si z těchto metadatových standardů vytváří tzv. metadatový aplikační profil. Koncept tohoto profilu je založen na ideji, že pro konkrétní kontext je nutno metadatové standardy lokalizovat a optimalizovat (49, s. 54). Metadatový aplikační profil je soubor metadatových elementů, které jsou vybrány z jednoho nebo více metadatových standardů a jsou spojeny do jednoho sloučeného schématu, který je uzpůsoben na míru funkčním požadavkům konkrétního užití, zatímco je zachována interoperabilita s původními standardy (55). Součástí profilu může být i vypracování vlastních metadatových elementů, v praxi digitalizátů knih to však není obvyklé.

Metadatový profil se může stát například národním standardem. Tak je tomu v případě ČR, kdy se metadatový profil vytvořený původně pro projekt NDK stal závazným standardem pro digitalizační projekty financované dotačním titulem VISK (56).

### 3.4.3 PREMIS

Standard PREMIS sám sebe označuje jako standard pro archivační metadata (*preservation metadata*), který „podporuje životaschopnost, reprodukovatelnost, srozumitelnost, autenticitu a identitu digitálních objektů v archivačním kontextu“ (52, s. 1). Slouží však nejen pro zápis archivačních informací, ale také interpretačních informací. Aktuální třetí verze PREMIS vyšla v roce 2015 (52). Standard neobsahuje pouze sadu elementů, ale také vlastní komplexní datový model, terminologický slovník a podrobný text vysvětlující logiku a možnosti užití standardu v archivu. Ve své sebedefinici klade standard PREMIS také důraz na to, aby jeho elementy byly implementovatelné, což podle něj znamená, že hodnoty většiny elementů musí být možné automatizovaně vyplňovat a zpracovávat archivem (54, s. 3). K takovému cíli ostatně směřují všechny zde uvedené metadatové standardy.

Datový model v PREMIS 3 definuje čtyři základní entity: objekt (*object*), činitel (*agent*), událost (*event*) a právní deklarace (*rights statement*). Objekt dále člení na čtyři úrovně: intelektuální entita (*intellectual entity*) je „jednotlivý intelektuální nebo umělecký výtvar (*creation*), který je považován za relevantní pro cílovou komunitu v kontextu digitální archivace“; reprezentace (*representation*) „množina souborů (včetně strukturálních metadat) potřebná pro úplnou reprodukci intelektuální entity“; soubor (*file*) „pojmenovaná a uspořádaná posloupnost bajtů, kterou dokáže rozeznat operační systém“ a která je v určitém formátu; bitový tok (*bitstream*) „data v rámci jednoho souboru, která mají smysluplné společné vlastnosti pro archivační účely“ (52, s. 8). Všechny úrovně (vyjma intelektuální entity) odpovídají pojmu „digitální objekt“ v modelu OAIS, přičemž reprezentace v PREMIS odpovídá pojmu „objekt CDO“. Intelektuální entita odpovídá informačnímu obsahu modelu OAIS s tím rozdílem, že v standardu PREMIS jde specificky o reprodukováný informační obsah (tj. který může vnímat člověk).

Intelektuální entitu je možné podle modelu PREMIS také dále specifikovat podle úrovní abstrakce popsaných ve známém knihovnickém modelu FRBR. Model FRBR stanovuje tyto čtyři úrovně: (*work*), vyjádření (*expression*), manifestace (*manifestation*) a exemplář (*unit*) (57).

PREMIS obsahuje elementy, které odpovídají všem typům archivačních informací. Klíčové jsou zejména elementy pro zápis provenienčních informací. V tomto ohledu PREMIS vhodně předepisuje logiku metadatového zápisu: „metadata, soubory, bitové toky a reprezentace uchovávané v archivu se popisují jako statické množiny bitů. Není možné změnit soubor (nebo bitový tok nebo reprezentaci); lze pouze vytvořit nový soubor (nebo bitový tok nebo reprezentaci), který se vztahuje k zdrojovému objektu.“ (52, s. 22). Tento vztah mezi novým a předchozím objektem definuje jako vztah odvození (*derivation relationship*), u něhož musí být zaznamenán specifický typ události, odlišný od událostí, které nevytvářejí nový objekt. Standard odlišuje dva typy odvození ze zdrojového digitálního objektu do nového objektu: replikace (*replication*) a transformace (*transformation*) (52, s. 19). Replikace znamená vytvoření digitální kopie, která je bitově identická se zdrojovým digitálním objektem (52, s. 272), transformace má za výsledek vytvoření jednoho nebo více digitálních objektů, které nejsou bitově identické se zdrojovým objektem (52, s. 273).

Pro strukturální interpretační informace slouží sekce elementů popisujících formát (název formátu; verze formátu; název formátového registru; identifikátor záznamu formátu v tomto registru; role registru). Pro podrobnější popis interpretačních informací je ve standardu PREMIS vyčleněna možnost vnořit externí schéma.<sup>34</sup> Pro digitalizáty knih je za tímto účelem užíván standard MIX. PREMIS obsahuje i sekci signifikantních vlastností, která však není v praxi příliš užívána. Problému signifikantních vlastností a dalším aspektům standardu PREMIS se budeme věnovat v Kapitolách 5-7.

Závěrem je třeba upozornit na to, že PREMIS 3.0 přinesl řadu změn, takže výše uvedené shrnutí neodpovídá tomu, které jsem prezentoval na několika místech ve své knize a které se vztahovalo k verzi 2.2 (1). V předchozí verzi standardu PREMIS byla intelektuální entita jednou z entit, nikoliv úrovní objektu. Nová verze standardu nově umožňuje také popis fyzických objektů. Naneštěstí při závěrečné redakci nedošlo k sjednocení definic obsažených v různých místech dokumentu, takže to nemusí být na první pohled zřejmé.<sup>35</sup>

---

<sup>34</sup> V rámci elementu „objectCharacteristicsExtension“.

<sup>35</sup> Srv. definici pojmu „reprezentace“ na straně 8 a straně 33 (pouze v druhém případě definice obsahuje údaj o tom, že tento pojem se nově vztahuje na digitální i fyzické objekty).

### 3.4.4 METS

Standard METS (50) slouží primárně jako metadata zaznamenávající informace o zabalení modelu OAIS (tedy o zabalení balíčků SIP, AIP a DIP). Především umožňuje vnoření dalších metadatových schémat pro popis archivačních a interpretačních informací (a tím jejich identifikaci). Dále obsahuje sekci určenou pro zápis provenienčních informací (formou vnoření externího schématu) a zápis některých interpretačních informací (např. o chování objektu). V praxi je METS užíván zejména pro první funkci (záznam informací o zabalení) a také jako datový formát. Jeho sekce strukturální mapa se využívá pro záznam informací o všech obrazových souborech a jejich posloupnosti. Tyto informace tedy tvoří strukturální interpretační informace, ale vlastní datovou součást digitalizátu knihy, bez níž by objekt CDO byl neúplný.

METS je možno užít v kombinaci se standardem PREMIS, přičemž lze zvolit několik způsobů implementace. Americké směrnice NISO doporučují zaznamenat PREMIS do sekce METS pro zápis provenienčních informací (30, s. 55). V praxi nalezneme různé způsoby. Například finská národní knihovna využívá METS v podobě jediného souboru pro celý digitalizát v balíčku AIP (58). Projekt NDK využívá více souborů METS (jeden pro bibliografická metadata, ostatní vytvářené zvlášť pro každý soubor obrazové matrice v archivním formátu) (56).

### 3.4.5 MIX

Název standardu MIX je zkratkou pro „Metadata for Images in XML Standard“. Tento standard je XML schéma, které je založeno na americké normě ANSI/NISO Z39.87-2006 (59). Podle vlastního popisu je účelem normy „standardizovaná sada metadatových elementů pro rastrová obrazová data“, přičemž tyto elementy „dokumentují digitální obrazová data vytvořená digitální fotografií nebo skenováním a též data, která byla pozměněna editováním nebo obrazovým převodem“ (59, s. 1). Standard MIX obsahuje elementy této normy, přidává několik dalších (např. rozděluje prostorové rozlišení do dvou elementů) a snižuje povinnost vyplnění elementů. Podle své vlastní definice MIX vznikl jako formát pro výměnu nebo uložení dat specifikovaných v uvedené normě NISO (54). Standard MIX se v praxi užívá pro záznam obrazových vlastností digitalizátů (tedy dalších typů interpretačních informací), a to jako externí schéma pro PREMIS.

Norma ANSI/NISO Z39.87-2006 uvádí, že není určena pro záznam provenience (59, s.1). Kupodivu to není tak docela pravda vzhledem k tomu, že jedna její sekce elementů („Change History“) je určena pro záznam provenienčních informací z doby produkce



(pro záznam generací dat vzniklých při vytváření finálních produkčních dat i užitých procesů), ale v praxi se za tímto účelem užívají spíše elementy standardu PREMIS, ačkoliv je PREMIS primárně určen pro záznamy operací v archivu, nikoliv pro digitalizaci.

### **3.5 Metadata v obrazových souborech**

Speciální oblastí je zabudování metadat do obrazových souborů. Široce rozšířeným a obrazovým průmyslem podporovaným standardem pro zabudovaná metadata je EXIF (60). Dva hlavní snímkové formáty (TIFF a RAW) podporují záznam metadat ve formátu EXIF. Tato metadata do souborů zapíše snímací zařízení (ať již skener, nebo fotoaparát) automaticky. Obsahují velké množství elementů popisujících mj. snímací zařízení (včetně sériového čísla), způsob snímání nebo obrazové nastavení. Formát JP2 záznam EXIF metadat neumožňuje. Umožňuje však vnořit libovolná metadata v XML (do strukturálního prvku „XML Box“, jehož volba je součástí profilu tohoto formátu). V praxi se do něj zapisují například bibliografické údaje (61).

## 4 Etapy životního cyklu digitalizátů knih

Pro popis životního cyklu digitalizátů lze využít normu ISO 14721 tak, že za uzlové body cyklu se bude považovat vytvoření balíčků SIP, AIP a DIP.

První etapou je produkce. Produkce končí vytvořením balíčku SIP a jeho dodáním do archivu. Za začátek produkce je třeba považovat již přípravu knihy k digitalizaci. Za etapu produkce je zodpovědný producent, za dodávání do archivu vkladatel, v našem případě obě funkce zastává stejný subjekt (jedna knihovna).

Druhou etapou je archivace. Archivace začíná vytvořením balíčku AIP z balíčku SIP (konkrétně vytvořením první verze balíčku AIP), který je následně uložen do funkčního celku Archivní úložiště a zde archivován (dlouhodobě uchováván). Archivace balíčku AIP znamená, že v závislosti na zastarávání technologického prostředí bude archiv muset postupně vytvářet nové verze balíčku AIP. Dílčí etapy (fáze) archivace lze stanovit jako vytvoření nové verze balíčku AIP. V daný okamžik je nejnovější verze balíčku AIP aktuální verzí, která se považuje za výstup aktuální implementace archivačních opatření. Aktuální verze balíčku AIP by tedy měla být taková, která zachovává původní informační obsah a jeho stanovené vlastnosti a současně je použitelná v aktuálním počítačovém prostředí. V současné praxi to znamená především, že objekt CDO je v (aktuálně vhodném) archivačním formátu, tj. který není ohrožen zastaráváním nebo uplatňováním patentů a má otevřenou dokumentaci. Balíček AIP první verze podle modelu OAIS může, ale nemusí být zachován (pro potřeby případného ověření zachování informačního obsahu v aktuální verzi balíčku AIP). O zachování verzí mimo první a aktuální verzi balíčku AIP norma OAIS nemluví a pravděpodobně nejsou tyto „meziverze“ z hlediska archivace podstatné.

Třetí etapou je zpřístupňování. Zpřístupňování začíná vytvořením balíčku DIP z balíčku AIP první verze. Následně mohou být vytvářeny další balíčky DIP v závislosti na změnách požadavků cílové komunity. Balíčky DIP by měly být vytvářeny z aktuální verze balíčku AIP. V praxi knihoven se zpřístupňování uskutečňuje tím, že balíčky DIP jsou nahrány do digitální knihovny, která je zpřístupňuje tak, že je zobrazuje v internetovém prohlížeči.

## 4.1 Produkce

### 4.1.1 Základní aspekty produkce

Pro produkci se zpravidla používá komplexní softwarový digitalizační systém, který řídí celý proces včetně evidence souborů (automatické pojmenovávání, indexace atd.). Tento systém zpravidla využívá další externí nástroje. Tato automatizace je nezbytná u masové produkce, například z hlediska dodržení posloupnosti kroků, řízení oprávnění a zabránění současně prováděných různých operací. Nejrozšířenějším digitalizačním systémem je komerční DocWorks, který využívá řada velkých knihoven.<sup>36</sup> Pro pojmenování souborů doporučují směrnice FADGI užití názvů, které jsou jedinečné v celém digitalizačním projektu (tedy nejen v balíčku SIP jednotlivého digitalizátu knihy), a následující pravidla (20, s. 79-81)

- název souboru musí končit tečkou, po níž následuje příslušná přípona souboru (zkratka formátu souboru);
- název souboru smí obsahovat pouze malá písmena (platí i pro přípony souborů);
- název souboru smí obsahovat maximálně 32 znaků;
- přípona musí obsahovat pouze tři znaky (např. „tif“, nikoliv „tiff“).

Důležitou otázkou jsou katalogizační záznamy tištěné předlohy. Směrnice MINERVA uvádějí: „každý fyzický objekt by měl být zkatalogizován před jeho digitalizací.“ (28, s. 25). Pokud katalogizační záznam existuje, měla by být provedena jeho kontrola ještě před tím, než se zahájí digitalizace. V praxi se míra této kontroly může lišit.<sup>37</sup> Pro identifikaci exempláře knihy se v současné praxi někdy užívá způsob, kdy se vytvoří sada čárových kódů, které se nalepí na list, jenž se do knihy vloží a poté odstraní.

### 4.1.2 Fáze snímání

Výstupem fáze snímání jsou původní snímky. Formát těchto snímků budeme nazývat snímkový formát. Obecně doporučeným postupem, který je v praxi bez problémů přejímán, je snímání do nekomprimovaného formátu. Užívají se dva typy snímkových formátů: TIFF pro skenování a formát rodiny RAW pro fotografování. Snímkový formát může být potřeba na snímacím zařízení zvolit, neboť tato zařízení často umožňují snímat do různých snímkových formátů, resp. formátových profilů; obrazové vlastnosti (prostorové rozlišení,

---

<sup>36</sup> <https://content-conversion.com/>

<sup>37</sup> Například ve finské národní knihovně se provádí podrobná kontrola, v NK ČR pouze kontrola základních údajů.

bitová hloubka, model RGB) se musejí vždy nastavit. Směrnice FADGI doporučují nastavit na skeneru snímání do barevného profilu Adobe RGB 1998 (pokud to snímací zařízení umožňuje) (20, s. 48). U fotoaparátu se musí nastavit snímání do formátu RAW, a případně vypočítat prostorové rozlišení, které se pak musí zapsat do souboru obrazu. Tento výpočet vyžaduje změření šířky a výšky tištěné knihy. V opačném případě nebude možné digitalizát vytisknout v původní velikosti.

Při digitalizaci se stránky obvykle snímají v původní posloupnosti knihy, ale užívají se i jiné postupy, například snímání nejprve všech levých stran knih, a poté všech pravých, či obráceně.<sup>38</sup> V obou případech se tato posloupnost převádí do digitální podoby zpravidla vhodně zvoleným systémem pojmenování souborů, takže je důležité, aby nedošlo k přeskočení stránky (při skenování může taková chyba nastat např., pokud jsou stránky slepené).

#### **4.1.3 Fáze zpracování obrazových dat**

Tato fáze zahrnuje zpracování původních snímků až do vytvoření finálních produkčních dat a může zahrnovat i vytváření prezentačních variant obrazových dat.

##### **4.1.3.1 Strukturální rekonstrukce a obrazové úpravy**

Fáze zpracování obrazových dat zahrnuje dva základní typy zpracování, které můžeme označit za strukturální rekonstrukci a obrazové úpravy. Strukturální rekonstrukcí zde myslíme úpravu snímků tak, aby co nejvíce odpovídaly stránkám tištěné knihy. Zahrnuje především ořezy a rekonstrukci posloupnosti stran. Rekonstrukce posloupnosti se provádí zpravidla automatizovaně (užitím systému pojmenovávání souborů pro celý digitalizační projekt), přičemž kontrolu posloupnosti je nutno provádět manuálně. Směrnice FADGI doporučují neprovádět žádné ořezy, s tím že „kolem celého dokumentu by měly být viditelné malé okraje.“ (20, s. 50). V současné praxi se ořezy někde provádějí (39), někde ne [(62) (34)].

Směrnice FADGI doporučují provádění menších obrazových úprav na původních snímcích: barevných korekcí, úprav tonality a doostření (20, s. 47). V praxi se na původních snímcích často žádné úpravy neprovádějí. Někdy jako projev přesvědčení, že tímto způsobem se zajistí nejvyšší možná kvalita (34), jindy čistě pragmatický přístup – pracnost takové operace (63). Za součást obrazových úprav lze považovat i převod do doporučeného

---

<sup>38</sup> Tento postup užívá švédská národní knihovna, která jej odůvodňuje zrychlením digitalizace (snímacím zařízením je zde fotoaparát) (34, s. 103).

barevného profilu. Směrnice FADGI doporučují uskutečnit převod do barevného profilu Adobe RGB 1998 při zpracování původních snímků, pokud tak nelze učinit již při nastavení skeneru (20, s. 48).

Zpracování se obecně doporučuje vykonávat na původních snímcích. Tak je tomu běžně při zpracování obrazů ve snímkovém formátu TIFF. Původní snímky ve formátu RAW se často zpracovávají v aplikacích firmy Adobe (34). Užití těchto aplikací znamená, že snímkový formát je automaticky aplikací Adobe převeden do formátu DNG. Ten úpravy ukládá do systémových souborů, nikoliv přímo do souboru ve formátu DNG. Úpravy se uloží až po exportu do formátu TIFF. Pracovat s formátem RAW zpravidla dokáží pouze aplikace výrobce fotoaparátu. Vlastním testem (užitím fotoaparátu Canon EOS 5D Mark II) bylo zjištěno, že aplikace firmy Canon dokáže ukládat do formátu Canon RAW změny (např. ořezy), a sice tak, že nastavené změny zapíše do strukturálních prvků formátu a úpravy se rovněž projeví až po exportu do formátu TIFF). Aplikace firmy Adobe i Canon umožňují při konverzi do formátu TIFF nastavit barevný profil Adobe RGB 1998. V případě zpracování snímkového formátu TIFF záleží na tom, jaké aplikace se užívají. Digitalizační projekty někdy využívají opensourcové nástroje, které většinou nedosahují možností nabízených profesionálními komerčními produkty, jako jsou programy firmy Adobe.

#### **4.1.3.2 Formátové konverze do archivačního formátu**

V současné praxi byly zjištěny následující varianty konverzních procesů od snímkového formátu do finálních produkčních dat v archivačním formátu. Tam, kde je snímkovým formátem TIFF (při skenování), ve většině případů zůstává archivačním formátem také TIFF, a tedy k žádné formátové konverzi při zpracování nedochází.<sup>39</sup> Méně častým případem je, že se provádí konverze do formátu JP2.<sup>40</sup>

Tam, kde je snímkovým formátem RAW (při fotografování), je první konverzí obvykle převod do konverzního meziformátu DNG a následuje konverze do archivačního formátu TIFF.<sup>41</sup> Druhou variantou je konverze do formátu TIFF a následně do formátu JP2.<sup>42</sup>

Pro převod do formátu JP2 z formátu TIFF se v praxi užívají různé nástroje (kodeky). Nejužívanějším je zřejmě kodek firmy Kakadu.<sup>43</sup> Jistý problém představuje, že jde

<sup>39</sup> Srv. (35). Tento postup užívá například Kongresová knihovna nebo francouzská národní knihovna (33).

<sup>40</sup> Matematicky bezeztrátovou kompresi pro archivní formát JP2 užívá například projekt NDK (39) a všechny současné projekty financované z podprogramu VISK 7 (48) nebo norská národní knihovna (38).

<sup>41</sup> Např. (34).

<sup>42</sup> Např. (117).

o komerční nástroj. Poslední dobou se rozšiřuje užití opensourcového nástroje Openjpeg (64). Kvalita jeho konverze byla zpočátku problematická, ale v nových verzích doznal výrazného zlepšení. V roce 2013 jej specialisté Britské národní knihovny (British Library) označili za nejlepší kodek pro kódování i dekódování do formátu JP2 (dalšími zkoumanými nástroji byl právě kodek Kakadu a Jasper) (65). Openjpeg je jedinou volbou pro organizace, které si nemohou dovolit zakoupení komerčního nástroje.<sup>44</sup> Pro převod do formátu JP2 se využívá formát TIFF jako konverzní meziformát vzhledem tomu, že kodeky neumějí přímý převod z formátu RAW.

#### **4.1.3.3 Vytváření prezentačních variant**

Někdy se již při digitalizaci vytvářejí prezentační varianty. Například švédská národní knihovna vytváří prezentační varianty z digitalizačního meziformátu DNG, na kterém nejprve provede úpravy (ořezy, barevné korekce (34). V Britské knihovně se prezentační varianty vytvářejí z obrazové matrice v archivačním formátu TIFF: obrázkové náhledy (GIF, JPEG a PNG), obrazy v nižším rozlišení (JPEG) a obrazy ve vysokém rozlišení (JPEG) (66, s. 10-12). V projektu NDK i v digitalizačních projektech podprogramu VISK7 se z formátu TIFF vytváří archivační formát (JP2 bezztrátová komprese) i prezentační meziformát (JP2 ztrátová komprese), z něhož pak digitální knihovna generuje prezentační formát JPEG (67), (48). V jiných případech se užívá jako archivační i prezentační formát jeden a tentýž formátový profil JP2 (37).

#### **4.1.3.4 Fáze vytváření a ukládání výstupu OCR**

Obecným pravidlem je, že proces OCR se musí provádět na produkčních datech po zpracování (po ořezech). Může se provádět na archivačním formátu, případně na prezentační variantě. Může být také vytvořen dočasný extrakční meziformát v komprimované podobě pro zlepšení výsledků procesu OCR. Mezi nejužívanější nástroje pro OCR patří ABBYY Fine Reader, který je komerčním produktem.<sup>45</sup> Nejznámějším opensourcovým nástrojem je patrně Tesseract.<sup>46</sup> Kvalita výstupů OCR mj. také závisí na slovnících, který software pro OCR užívá.

---

<sup>43</sup> Wellcome Library (37), projekt NDK projekt (39), přehled dalších viz (40).

<sup>44</sup> V digitalizačních projektech financovaných z podprogramu VISK7, kde je povinnost vytvářet data v archivačním formátu JP2, knihovny většinou využívají kodek openjpeg.

<sup>45</sup> <http://www.abbyy.cz/>

<sup>46</sup> <https://github.com/tesseract-ocr/> patří

#### 4.1.4 Fáze vytváření metadat

Specifikaci toho, jaká metadata mají být vytvořena, stanovuje metadatový aplikační profil pro produkci. Aplikační profil se v současné praxi vesměs vytváří užitím uvedených metadatových standardů (viz oddíl 3.4). Současná praxe se vyznačuje snahou zachytit při digitalizaci co nejvíce technických informací o tomto procesu (tj. o užitých snímacích zařízeních, konverzních nástrojích, procesech zpracování apod.), a to automatizovaným způsobem a co nejdříve vzniku digitalizátu.<sup>47</sup>

Užité metadatové standardy samy o sobě určují okruh informací, které lze zaznamenat, a stanovují poměrně přesně vymezené elementy. Pro některé elementy doporučují využití řízených slovníků. Ty jsou důležité zejména pro standard PREMIS. Ten sám často odkazuje na řízené slovníky Kongresové knihovny, které jsou však poměrně omezené a umožňují standardizaci pouze několika typů obsahu. Optimálním přístupem, a to nejen pro fázi produkce, ale zejména archivace, je, aby si organizace vytvořila vlastní řízené slovníky pro podrobnější specifikaci možného obsahu metadatových elementů, ale pro fázi produkce není takový postup běžným.

V praxi se nesetkáváme s tím, že by jednotlivé organizace prováděly výzkum, jaké všechny informace lze v průběhu digitalizace zachycovat. Příprava digitalizačního projektu z metadatové stránky znamená obvykle prostudování dostupných standardů a následná volba takových elementů, které budou začleněny do profilu a jejichž hodnoty se budou v průběhu digitalizace zaznamenávat. Metadatové standardy jsou tak nejen vodítkem pro organizace, ale zároveň vymezují maximum zaznamenateľných údajů (vyjma případů, kdy si organizace vytváří vlastní schémata, což není běžný jev).

Fáze vytváření metadat může zasahovat do všech časových okamžiků digitalizace, proto jí nelze přiřadit konkrétní pořadí v etapě produkce. Informace o konkrétním způsobu a okamžiku vyplňování v digitalizaci lze obtížně zjistit. Do značné míry jde o otázku nastavení digitalizačního systému, který většinu operací souvisejících s metadaty provádí automatizovaně a pracovníci organizace často nemají bližší představu o tomto procesu, přičemž jeho dokumentace nebývá veřejná.

Některá metadata lze vytvářet již před zahájením vlastní digitalizace knihy (zejména bibliografické informace o předloze, na které digitalizační postup nemá vliv). Některá metadata z logiky věci nelze vytvořit dříve než ve fázi vytváření balíčku SIP (např. informace

---

<sup>47</sup> Srv. např. (118, s. 35-39).

o zabalení balíčku SIP). Jak již bylo uvedeno, optimální kontrola struktury digitalizátu knihy musí být provedena manuálně a případné opravy budou zasahovat do metadat. Pro ostatní okruhy informací je obtížné určit, v jaké konkrétní fázi se vytvářejí. V řadě případů se uchovávají všechny generace obrazových dat a mažou se až po úspěšném dodání balíčku SIP do archivu, takže z tohoto hlediska je technicky možné vytvářet metadata o starších i novějších generacích ve stejný okamžik.

Metadata se v současné praxi získávají následujícími způsoby: převodem z jiných systémů (např. katalogizačního systému), z přednastavených hodnot (např. informace o snímacím zařízení jedné digitalizační linky), extrakcí ze souborů a lidskými pracovníky. Pro získávání metadat ze souborů existují dva typy nástrojů: nástroj pro formátovou identifikaci a extraktor metadat.

Značnou část informací ze souboru je možno získat extrakcí metadat EXIF, které snímací zařízení zabudovává do snímkových formátů TIFF nebo RAW jako standardní součást snímání. Při následných formátových konverzích však mohou být tyto informace ztraceny (např. formát JP2 nebo PNG zabudování metadat EXIF nepodporují). Směrnice FADGI vydala doporučení, jaké elementy EXIF metadat je vhodné zaznamenat (68). Velká výhoda těchto metadat spočívá v tom, že je z nich možno získat například informace o snímacím zařízení. Pokud jsou tyto informace zaznamenávány do metadat z přednastavených hodnot uložených v digitalizačním systému, hrozí vždy lidská chyba (např. při výměně skeneru se provede změna přednastavených hodnot).

Specifickou oblastí je zabudování metadat do obrazových souborů. V praxi se tak děje v některých projektech využívajících formát JP2. Kongresová knihovna do metadat formátu JP2 zabudovává bibliografická metadata včetně identifikátoru LCCN ve schématu DC (61, s. 12), Wellcome Library navíc též identifikátor UUID (37, s. 11).

#### **4.1.5 Fáze vytváření balíčku SIP a jeho dodávání do archivu**

Tato fáze je poslední fází produkce. Znamená uložení všech částí digitalizátu knih a metadat do jednoho balíčku SIP.

Balíček SIP je v současné praxi obvykle strukturován do podoby adresáře s podadresáři obsahujícími odlišné typy datových objektů (např. soubory v archivačním formátu, soubory prezentační varianty, soubory s výstupy OCR). Za účelem dodávání do archivu je adresář s celým obsahem zabalen do jednoho souboru v balíčkovacím formátu, často je jím ZIP nebo TAR. Specificky pro digitální archivaci však byly vyvinuty dva



balíčkovací formáty: BagIt File Packaging Format (standard IETF)<sup>48</sup> a WARC (standard ISO 28500). Formát BagIt využívá řada amerických knihoven. Formát WARC je užíván zejména ve webarchivech (pro zabalení sklizeného webového obsahu), nicméně například italská národní knihovna jej využívá pro zabalení elektronických publikací.

Dohoda o dodávání dat by měla stanovit datový model (vymezení toho, jaké soubory nebo jejich části tvoří objekt CDO, jaká metadata zaznamenávající interpretační informace a jaká metadata obsahující archivační informace), jehož součástí by měl být datový slovník (typologie dat z hlediska formátu nebo užití). V rámci jedné organizace může mít dohoda o dodávání dat podobu specifikace obrazových dat a metadat, která je součástí digitalizačního projektu. Nicméně obecně je vhodné, aby měla oficiální podobu, mj. vzhledem k tomu, že digitalizaci a archivaci v praxi obvykle zajišťují jiné organizační útvary téže knihovny. Součástí dohody by měl být i způsob řešení oprav chyb a bez oficiální podoby schválené managementem archivu (tedy vedením knihovny) mohou nastat zbytečné problémy vyplývající z nejasně definovaných povinností, resp. jejich vymahatelnosti.<sup>49</sup>

## **4.2 Archivace**

Úvodem této podkapitoly je nutno poznamenat, že většina národních knihoven a dalších paměťových institucí se nachází spíše ve fázi budování archivů OAIS. Nelze tedy konstatovat, že by všechny knihovny naplňovaly všechny stanovené činnosti funkčních modelu OAIS. Dosud mi není známa jediná národní knihovna světa, která by pro svůj archiv získala certifikát důvěryhodného archivu podle normy 16363. Předchozí certifikační standard (TRAC) získalo jen několik organizací.

### **4.2.1 Technické řešení archivace**

Organizace obvykle užívají různé softwarové systémy naplňující alespoň částečně činnosti funkčních celků archivu OAIS, přičemž některé činnosti nutně vykonávají lidé jako kurátoři (obsluhující tyto systémy) a metodici (sledující technologický vývoj, vytvářející standardy apod.). Francouzská národní knihovna již více jak deset let buduje vlastní softwarový systém SPAR, který byl nedávno obohacen o další modul určený pro formátovou strategii (69). Portugalská národní knihovna vyvíjí opensourcový systém RODA. Z komerčních systémů jsou nejužívanějšími Rosetta firmy ExLibris a Preservica. Ani tyto

---

<sup>48</sup> <https://tools.ietf.org/html/draft-kunze-bagit-14>

<sup>49</sup> Ve francouzské národní knihovně například taková oficiální dohoda mezi různými útvary knihovny existuje.

systemy se však neobjedou bez začlenění některých dalších nástrojů, z nichž nejdůležitější jsou specifické nástroje vyvinuté pro potřeby digitální archivace. Nejrozšířenější opensourcový systém Archivematica nabízí mj. vytvoření metadat v METS a PREMIS, určité typy normalizací a vytvoření balíčku AIP zabaleného ve formátu BagIt.<sup>50</sup> Tento systém je de facto celý založen na spojení různých specializovaných nástrojů pro digitální archivaci.

Obecně lze říci, že v současné praxi má většina knihoven vyvinutý přinejmenším systém pro funkční prvek Archivní úložiště. Běžným způsobem je ukládání balíčků AIP ve třech identických kopiích a užití profesionálních úložných systémů pro zajištění neporušenosti dat.

#### **4.2.2 Opatření vykonávaná na balíčku SIP v archivu**

Vzhledem k tomu, že v současné praxi jsou vytvářena finální produkční data zpravidla již v archivačním formátu (TIFF nebo JP2), při příjmu dat není třeba nevykonávat formátovou normalizaci do archivačního formátu. Základními funkčními prvky jsou Kontrola kvality (neporušenosti souborů na základě užitého digitálního otisku) a Kontrola vstupních dat z hlediska jejich souladu s metadatovým aplikačním profilem, formátovým profilem a případně i požadovanými obrazovými vlastnostmi. Za tímto účelem se využívají již zmíněné specializované nástroje pro digitální archivaci. Blíže o nich pojednává oddíl 7.5.

Specifickou oblastí je kontrola metadat. Pro ni neexistují rozšířené nástroje, a to z tohoto důvodu, že knihovny užívají specifické metadatové aplikační profily. Vývoj takového nástroje je velmi náročná činnost, jak se zjistilo i v projektu NK ČR pro vytvoření komplexního validátoru (70). V praxi projektu NDK se také ukázalo, že nejčastější činností archivu jsou různé opravy digitalizovaných dokumentů.

#### **4.2.3 Opatření vykonávaná v průběhu uchovávání**

Formátová konverze je v současné praxi hlavním zvažovaným archivačním opatřením proti zastarávání formátu. Reálně však k žádnému zastarávání dvou hlavních formátů pro obrazová data v balíčku AIP (TIFF a JP2) nedochází. Formát TIFF verze 6 se již více jak 25 let těší vysoké podpoře i užívanosti. JP2 je mladší formát, jehož technologická podpora dokonce narůstá.

V současné praxi dochází k jevu, který je z hlediska teorie digitální archivace překvapivý. Některé velké organizace (např. francouzská nebo britská národní knihovna)

---

<sup>50</sup> Srv. novější přehled (113, s. 29-30).

se rozhodly převést svoje digitalizované dokumenty z archivačního formátu TIFF do formátu JP2 ve ztrátové kompresi z důvodu úspory místa (65), (69, s. 57). K podobnému kroku již před několika lety přistoupila Wellcome Library, která tento převod prezentovala jako „kompromis mezi cenou za zajištění dlouhodobého digitálního uchovávání a užitými standardy pro obrazová data. Očekává se, že užitím formátu JP2 budou celkové požadavky na uložení udrženy ve výši, která představuje přijatelný kompromis mezi náklady uložení a obrazovou kvalitou“ (37, s. 3). Sledujeme tedy jev, kdy dochází k formátové konverzi v archivu nikoliv z důvodu zastarávání formátu, ale z důvodu úspory úložných kapacit. Ačkoliv toto opatření zůstává archivačním opatřením (řeší riziko, které nevyplývá ze změn technologií, ale fungování organizace jako takové), jde o nutný kompromis, který ukazuje na skutečnost, jak se odlišuje teorie a praxe digitální archivace.

#### **4.2.4 Metadata pro archivaci**

Základem pro metadatový popis při archivaci je PREMIS. Ten je sice užíván, jak bylo uvedeno výše, i při digitalizaci, ale jeho primárním účelem je sloužit pro potřeby archivu. V praxi se také hojně využívá, zejména v americkém prostředí, jak ukazuje registr metadatových archivačních profilů.<sup>51</sup> Využívá jej i softwarový systém Archivematica. Standard PREMIS pro uchovávání v archivu předepisuje, aby každý soubor nesl jedinečný identifikátor v rámci celého archivu, a každá událost odvození musí být zaznamenána. Komerční produkty Preservica a Rosetta používají vlastní metadatový standard, který je proprietární.

### **4.3 Zpřístupňování**

Etapa zpřístupňování je klíčová z hlediska cílové komunity archivu. V současné praxi je tato cílová komunita ve většině případů (týká se národních knihoven, ale i řady dalších knihoven, s možnou výjimkou specializovaných vědeckých knihoven určených pro výzkumníky) určená obecně: je jí každý čtenář dané země.

Hlavní formou zpřístupnění současné praxe je prezentace digitalizované knihy v digitální knihovně. Digitální knihovnou zde rozumíme komplexní systém, který obsahuje webové rozhraní pro prezentaci digitalizátu knih čtenářům, úložný systém a další komponenty, včetně případného obrázkového serveru, ze kterého se z prezentačního meziformátu generují výsledné obrázky v prezentačním formátu. Základní technické

---

<sup>51</sup> <https://www.loc.gov/standards/premis/registry/>

požadavky, předpokládané na straně čtenáře (tj. takové, které jsou součástí jeho znalostní základny), jsou internetový prohlížeč a běžný počítač.

Takto koncipovaná digitální knihovna může být pojímána jako součást archivu OAIS (jako funkční celek Zpřístupnění). Do digitální knihovny se v tomto případě dodávají balíčky DIP vytvořené z balíčků AIP uložených ve funkčním celku Zpřístupnění. Tyto balíčky DIP se následně ukládají v úložišti digitální knihovny, které by v modelu OAIS odpovídal funkční prvek Vytvoření balíčku DIP, který zahrnuje dočasné úložiště. Konceptu dočasného úložiště v současné praxi také reálně odpovídá úložný systém digitální knihovny, v tom smyslu, že prezentační formáty (případně i prezentační meziformáty) se uchovávají jen po dobu, než budou nahrazeny jinými, vhodnějšími pro zpřístupnění cílové komunity. Prezentační formáty a meziformáty se v současné praxi ostatně nepovažují za předmět dlouhodobého uchovávání. V případě potřeby se vytvoří z obrazových matic v archivačním formátu uloženým v balíčku AIP. Z tohoto důvodu je úkolem digitální knihovny uchovávat prezentační varianty v původní datové podobě, tedy zaručit jejich neporušenost z hlediska neměnnosti bitů. Digitální knihovna jako část archivu OAIS pak balíčky DIP dále zpracuje, a to tak, aby byly prezentovány uživatelům jako vizuální náhledy jednotlivých stránek. V tomto smyslu tedy digitální knihovna funguje jako zpřístupňovací software (který spadá do kategorie ostatních interpretačních informací modelu OAIS).

V případech současné praxe, kdy se balíček SIP dodává do archivu a současně i do digitální knihovny, však nelze hovořit o tom, že by archiv a digitální knihovna tvořily jeden celek (podle modelu OAIS musí být balíček DIP vytvářen z balíčku AIP).<sup>52</sup> Zde digitální knihovna tvoří odlišný systém. Z hlediska řízení životního cyklu digitalizátů však přesto musí knihovna (jako jedna a tatáž organizace spravující tento cyklus) zajistit korelaci mezi obrazy v archivačních formátech uložených v archivu a jejich prezentačními variantami v digitální knihovně; například to, aby se v případě změn digitalizátů tyto změny projevíly v datech archivního formátu i prezentační variantě.

#### **4.3.1 Způsoby zpřístupňování v digitální knihovně**

Běžným způsobem zpřístupnění hlavní prezentační varianty je zobrazení digitalizátu knihy ve formě obrazů jednotlivých stránek, kterými je možno listovat. Alternativní digitální knihovna Moravské zemské knihovny nabízí možnost zobrazení knihy jako dvojstránky.<sup>53</sup>

---

<sup>52</sup> Jako je tomu v projektu NDK (67).

<sup>53</sup> <http://www.digitalniknihovna.cz/mzk/>

V digitální knihovně má digitalizát knihy vyčleněnu úvodní stránku s URL adresou, ze které jsou přístupné (přímým odkazem nebo řetězcem odkazů) všechny digitalizované stránky a další části knihy. Na tuto stránku také zpravidla odkazují užitím daného URL katalogy knihoven.

Prezentační formát může být uložen v úložném systému digitální knihovny. Při prezentaci může být nabízeno více prezentačních variant obrázků. Druhým způsobem je, že prezentační formát je za chodu (během prohlížení čtenářem) automaticky generován z prezentačního meziformátu uloženého v obrázkovém serveru digitální knihovny. Tento druhý způsob je v současnosti běžným způsobem zobrazení v případě, že je užit formát JP2 jako prezentační meziformát, z něhož se generují prezentační varianty ve formátu JPEG (37, s. 11).

Digitální knihovny někdy jako doplňkové způsoby zpřístupňování užívají například možnost stažení celého digitalizátu v jednom souboru (zpravidla ve formátu PDF). Někdy je nabízena možnost automatického vygenerování PDF za chodu, celého digitalizátu nebo jen jeho části. Někdy mají tato PDF výrazně nižší kvalitu než hlavní prezentační formát.<sup>54</sup>

Vzhledem k tomu, že technologie, která je široce dostupná cílové komunitě, se v poslední době rozšířila o chytré mobilní telefony, některé knihovny vyvíjejí jako doplňkovou službu mobilní aplikace sloužící pro jednodušší prohlížení digitální knihovny v chytrých telefonech (například digitální knihovna Gallica). Tyto telefony sice obsahují internetový prohlížeč, ale běžnou a uživateli očekávanou praxí je nabídka specializovaných mobilních aplikací, které zpřístupňují internetový obsah z konkrétní služby.

#### **4.3.2 Metadata v digitální knihovně**

Běžnou praxí od počátků digitálních knihoven je prezentovat spolu se zobrazenou knihou také přidružená metadata. Ta většinou obsahují pouze bibliografické údaje (identifikační informace podle modelu OAIS), někdy obohacené o údaj o projektu (pokud projekt nemá samostatné stránky). Nenalezl jsem případ, že by byla spolu s digitalizátem vystavována také archivační metadata. Zobrazená metadata s URL adresou úvodní stránky jsou pak základním zdrojem citování. URL adresa tvoří hlavní aspekt existence digitálního dokumentu v internetové síti. Pro citování i vyhledávání je však klíčovou novou službou systém perzistentní identifikace nezávislý na URL adrese, který dokáže přesměřovat na aktuální funkční URL adresu. Tuto službu blíže popisuje oddíl 9.1.

---

<sup>54</sup> Finská národní knihovna například nabízí PDF v nižší obrazové kvalitě a černobílé variantě (58).

## 5 Koncept autenticity digitálních dokumentů

Pátá kapitola přináší přehled vybraných konceptualizací a tematizací autenticity, které jsou relevantní pro téma této disertační práce.<sup>55</sup> Kritériem relevantnosti je, aby předmětem zkoumaných zdrojů byl datový objekt (tj. jednotlivina), který je považován za součást kulturního dědictví, případně je i dokumentem (tj. takovým datovým objektem, který je institucionalizovaně vytvářen nebo spravován).<sup>56</sup> Tematizací autenticity se rozumí jakékoliv odborné pojednání o autenticitě; konceptualizací autenticity se rozumí soustavná tematizace, která o autenticitě pojednává komplexně (tj. ze všech aspektů v kontextu, pro nějž byla vytvořena).

S konceptem autenticity se setkáme v řadě nejrozumnějších disciplín, převážně společenskovední a humanitní oblasti. Zatímco digitální informace se staly společenským fenoménem od konce dvacátého století, pojem autenticity a jemu přisuzovaný velký společenský význam má mnohem hlubší kořeny.

### 5.1 Historie pojetí autenticity dokumentu

Podle Luciany Durantiové sahá historie autenticity (ve smyslu autenticity dokumentu) až do starořímské právní teorie. Z ní byla v 17. století přejata do klasické diplomatiky a odtud později do archivní vědy (71). Klasická diplomatika vznikla původně jako disciplína zjišťující autenticitu starých úředních dokumentů z hlediska právních otázek. Toto pojetí autenticity bylo v 18. století vyučováno na evropských právních fakultách. V 19. století se přeměnila v pomocnou vědu historickou a jejím předmětem se stalo posuzování středověkých dokumentů jako historiografických pramenů (72). Archivní věda i klasická diplomatika se zabývají autenticitou dokumentů, přičemž předmětem klasické diplomatiky je jednotlivý dokument a předmětem archivní vědy sbírky (agregace dokumentů).

První významná konceptualizace autenticity digitálních dokumentů byla vytvořena v projektech vedených právě archivními teoretiky. Jedná se zejména o projekt InterPARES (1991-2001), který se věnoval stanovení požadavků pro udržování a posuzování autenticity elektronickým archiválií (73). Z tohoto důvodu následuje za tímto úvodem podkapitola

---

<sup>55</sup> Všechny zdroje uvedené v této kapitole jsou (až na jednu výjimku) psané v angličtině. Jelikož je v případě výrazu „autenticita“ anglický ekvivalent vždy tentýž („authenticity“), není uváděn v závorce u žádného zdroje.

<sup>56</sup> To nám umožňuje vyhnout se tematizacím, v nichž je předmětem autenticity obecnina (což je celá řada filosofických prací).

věnovaná přehledu konceptualizace digitálních dokumentů v archivnictví. Následuje podkapitola analyzující tematizaci autenticity v normě ISO 14721:2012 (10), která představuje základní (velmi obecný) rámec pro autenticitu digitálních dokumentů, jehož zdrojem jsou nepochybně právě výstupy archivářských projektů. Po ní je uvedena podkapitola věnovaná standardu PREMIS jako jednoho z hlavních praktických nástrojů pro podporu autenticity. Následuje podkapitola věnovaná tematizaci autenticity v knihovnictví, tj. ve zdrojích, které buď přímo vznikly v oblasti knihovnictví, nebo jsou v této oblasti hojně užívány. V oblasti knihovnictví dosud neexistuje konceptualizace autenticity digitalizátů knih v celém jejich životním cyklu, ani digitálních dokumentů obecně. Vybrané zdroje, které tvoří především hlavní standardy a směrnice v oblasti digitalizace a správy digitálních knihovních fondů, se zabývají autenticitou pouze jako jedním z mnoha dílčích témat. Následující podkapitola přináší přehled vybraných zdrojů jiných oblastí a tvoří doplněk k výše uvedeným přehledům.

## 5.2 Autenticita v oblasti archivnictví

Z disciplín, mezi jejichž předměty zájmu patří datové objekty digitálního dědictví, reagovala na potřebu nové konceptualizace životního cyklu dokumentů v souvislosti s prudkým rozvojem digitálních technologií a internetizace společnosti nejrychleji archivní věda. Již v roce 1994 byl zahájen tříletý severoamerický výzkumný projekt Preservation of the Integrity of Electronic Records, na který pak navázal mezinárodní projekt InterPARES (International Research on Permanent Authentic Records in Electronic Systems). Předmětem výzkumu těchto projektů byly archivní dokumenty spravované elektronickými systémy organizací. Archivní dokument (*record*) zde byl pojat jako jakýkoliv dokument, který byl vytvořen (nebo přijat) a uložen v průběhu praktické činnosti původce jako nástroj nebo vedlejší produkt této činnosti; elektronický archivní dokument (*elektronic record*) jako takový „archivní dokument, který je uložen a využíván v elektronické formě bez ohledu na původní formu, ve které byl vytvořen nebo přijat.“ (19, s. 15).<sup>57</sup>

Předmětem zkoumání a doporučení obou projektů byly především archivní dokumenty těch původců, kterými jsou státní organizace. První projekt se zaměřil na aktivní archivní

---

<sup>57</sup> V české legislativě odpovídá anglickému výrazu „record“ termín „dokument“. Tento český termín však není v této disertační práci užíván, vzhledem k tomu, že pojem dokument je zde užíván v širším významu, tedy jako jakýkoliv datový objekt vytvářený nebo spravovaný (uchovávaný, zpřístupňovaný) nějakou organizací (tj. digitalizovaná kniha je rovněž dokumentem).

dokumenty (tj. dokumenty, které jsou stále využívány v běžném chodu organizace), druhý na neaktivní archivní dokumenty (tj. dokumenty, které již původce pro svoji činnost nevyužívá, ale které musejí být dlouhodobě uchovávány např. z legislativních důvodů), tedy na archiválie.

### 5.2.1 Projekt Preservation of the Integrity of Electronic Records

Tento projekt vedla v letech 1994-1997 kanadská Univerzita Britské Kolumbie (University of British Columbia) ve spolupráci s archiváři amerického Ministerstva obrany USA. Hlavním výzkumníkem projektu byla kanadská teoretička archivní vědy Luciana Durantiová. Z cílů, které si projekt vytyčil, je pro téma dizertační práce relevantní cíl „určit konceptuální požadavky pro zaručení spolehlivosti a autenticity záznamů v elektronických systémech“ (74, s. 47). Spolehlivost (*reliability*) a autenticita jsou klíčové pojmy tohoto projektu, jejichž společným nadřazeným pojmem je integrita (*integrity*).

Výzkumný tým vytvořil následující definice. Spolehlivý archivní dokument (*reliable record*) je takový, se kterým „lze zacházet jako s faktem samotným, to jest jako s entitou, jejímž je dokladem (*evidence*).“ (75, s. 6); jeho spolehlivost je jeho důvěryhodnost (*trustworthiness*) jakožto dokladu (*evidence*), tj. jeho schopnost vypovídat o skutečnostech, o kterých pojednává.“ (74, s. 54). Archivní dokument je „autentickým, pokud je takovým dokumentem, za který se prohlašuje (*it claims to be*).“ (75, s. 7). Autenticita archivního dokumentu se vztahuje k režimu, formě a stavu přenosu archivního záznamu a rovněž ke způsobu jeho uchovávání a správy (*custody*). Ochrana a záruka autenticity dokumentu spočívají v zavedení metod, které zajistí, aby archivní dokument po svém vzniku nebyl pozměněn nebo zfalšován (74, s. 56).

Spolehlivost archivního dokumentu a způsoby jejího zaručení jsou podle týmu spojeny výhradně s procesem vytváření archivního záznamu (otázky míry jeho úplnosti a řízení postupu jeho vytváření). Autenticita archivního dokumentu znamená, že dokument je právě tak spolehlivý, jako byl spolehlivý v době svého vzniku: jinými slovy „pouze zaručuje, že archivní dokument není výsledkem manipulace, nahrazení nebo zfalšování, ke kterému by mohlo dojít po dokončení procesu jeho vytvoření, a tedy že [archivní dokument] je tím, za co se vydává (*what it purports to be*).“ (75, s. 7).

Pro zkoumání otázek autenticity vyšel projektový tým z následující logiky: aby bylo možno stanovit požadavky pro zachování a posuzování autenticity elektronického archivního dokumentu, je potřeba nejprve vytvořit model tohoto typu dokumentu. Za tímto účelem tým převzal koncepty klasické diplomatiky a vytvořil univerzální model archivního dokumentu,



skládají se z následujících základních složek (*constituents*): 1) obsah (*content*) - sdělení, k jehož přenosu má archivní záznam sloužit; 2) médium (*medium*) – fyzický nosič sdělení; 3) forma (*form*) – pravidla reprezentace, která umožňují přenos sdělení; 4) osoby (*persons*) – entity, které užitím archivního dokumentu jednají; 5) úkon (*action*) – projev vůle, jenž je příčinou vzniku archivního dokumentu jako prostředku pro vytváření, udržování, změnu nebo ukončování situací; 6) kontext (*context*) – právně-správní rámec, ve kterém se úkon odehrává; 7) archivní vazba (*archival bond*) – vztah, jenž propojuje archivní dokument se souvisejícími dokumenty stejné činnosti.

Užitím tohoto modelu v praxi elektronických systémů pak tým dospěl k závěru, že jak tradiční (tj. neelektronické), tak elektronické archivní dokumenty jsou tvořeny stejnými základními složkami. Rozdíl však spočívá v tom, že jednotlivé základní složky elektronických archivních dokumentů nejsou neoddělitelně spojeny (zatímco složky tradičních dokumentů tvoří neodlučný celek) a v elektronických systémech mohou být spravovány odděleně. Základním doporučením týmu pro zachování autenticity elektronických archivních dokumentů je zajistit, aby všechny tyto složky podléhaly jednotně řízené správě (74, s. 49).

Tým dále odlišil dvě základní fáze řízení životního cyklu dokumentů. První fází je řízení aktivních archivních dokumentů, které zahrnuje dva aspekty: a) vytváření spolehlivých elektronických archivních dokumentů, b) následné udržování autenticity těchto dokumentů. Za řízení této fáze by měl být odpovědný původce (tj. organizace, která dokumenty vytvořila a dále je využívá pro svoji činnost). Druhá fáze je správa elektronických archiválií (tj. těch neaktivních elektronických dokumentů, které byly vybrány k dlouhodobému uchovávání). Archiválie jsou uchovávány za jinými a širšími účely, než pro které byly původně vytvořeny. Za tuto fázi by měla mít odpovědnost „neutrální třetí strana“, buď samostatná specializovaná organizace (tedy tradiční archiv), nebo specializovaný útvar organizace původce (74, s. 60). Odpovědnost za správu archiválií se vztahuje již pouze na autenticitu, nikoliv spolehlivost. Pro řízení celého životního cyklu archivního dokumentu je nutno dodržet klasické archivářské pravidlo, jímž je „nepřerušená linie fyzické správy“ (74, s. 57).

Při srovnání tradičního uchovávání a digitální archivace přišel tým se závěrem, že základní archivační přístupy jsou v obou případech stejné: replikace (*copying*)<sup>58</sup> a konverze (*migration*) dokumentů. Replikace je vytvoření „úplné kopie obsahových i formálních prvků

---

<sup>58</sup> Durantiová v pozdějším článku, ve kterém prezentuje výstupy tohoto projektu, užívá odlišný výraz: „refreshing“ (71, s. 47).

archivního dokumentu“ (74, s. 50). Uváděnými příklady jsou tzv. „imitující kopie“ (tj. kopie dokumentů, které napodobují nejen obsah archivního dokumentu, ale i jeho formální znaky, např. rozložení textu - např. mikrofilm) nebo přesun bitů z jedné magnetické pásky na druhou. Konverze je „kopie obsahu dokumentu se změnami v konfiguraci a struktuře“ (74, s. 50). Uváděnými příklady jsou přepis tradičního dokumentu nebo převod elektronického dokumentu z jedné databáze do druhé mající jinou konfiguraci. Současně však výzkumný tým dodává, že při tradičním uchovávání se většinou zachovávaly i původní dokumenty, a tak bylo možno autenticitu výsledné kopie dokumentu ověřit srovnáním s originálem nebo s předchozími kopiemi; v digitálním světě, vzhledem ke křehkosti nosičů a zastarávání technologií, přežívají již jen kopie dokumentů (71, s. 47).

Pro digitální archivaci podle týmu platí, že kopie dokumentů vzniklé replikací nepředstavují větší problém, protože výslednou kopii dokumentu lze považovat za identickou (bitově totožnou) s původním dokumentem (liší se pouze nosič), a to z toho důvodu, že elektronický nosič není významotvorný prvek (nosič tradičního dokumentu je významotvorný) (74, s. 49). Problematická je tak především konverze, neboť znamená podstatnou změnu fyzické formy. Pro konverzi v první fázi životního cyklu se podle týmu i na elektronické dokumenty vztahuje klasický princip archivní vědy: dokud se organizace při své činnosti nepřestává spoléhat na kopie dokumentů vzniklé konverzí, toto spoléhání samo o sobě činí kopie dokumentů autentickými (74, s. 50). Základní záruku autenticity v této fázi tedy tým spatřuje v pragmatismu organizace.<sup>59</sup> Výzkumný tým vytvořil několik obecných doporučení pro takovéto organizace, mj. zabudovat procesní pravidla do celého systému správy elektronických dokumentů, vytvořit postupy pro vytváření dokumentů a zavést kontrolu procesů vytváření, správy a uchovávání napříč celou organizací (74, s. 58-59).

---

<sup>59</sup> Aby mohla organizace (jakožto původce) vykonávat svoji činnost, „má přímý zájem na vytváření a správě spolehlivých a autentických archivních dokumentů.“ (74, s. 50).

Hlavní výstupy tohoto projektu lze tedy shrnout do následujících závěrů:

- a) tentýž model archivního dokumentu lze užít pro tradiční i elektronické dokumenty, a to s tím rozdílem, že základní součásti elektronického dokumentu existují odděleně a pro udržení autenticity vyžadují jednotnou správu
- b) základní přístupy tradičního uchovávání a digitální archivace jsou stejné (replikace, konverze);
- c) princip, že spoléhání se organizace na kopie dokumentů činí tyto kopie autentickými, je platný i pro elektronické dokumenty (ve fázi jejich užívání původcem).

Durantiová dále ve své teoretické studii, která byla součástí výstupů projektu, uvedla, že nejdůležitějším úkolem archivnictví v digitálním světě je stanovit postupy pro vytváření spolehlivých elektronických dokumentů (tj. např. odpovídajících právním požadavkům): uchovávání autentických, ale nespolehlivých dokumentů postrádá smysl (75).

### **5.2.2 Projekt InterPARES**

Projekt InterPARES (International Research on Permanent Authentic Records in Electronic Systems) probíhal v letech 1999-2001 v návaznosti na projekt Preservation of the Integrity of Electronic Records. Vedoucím výzkumného týmu, tvořeného odborníky z několika zemí světa, byla opět Luciana Durantiová. Projekt InterPARES se věnoval autenticitě archiválií (tedy neaktivních archivních dokumentů, které byly vybrány k dlouhodobému uchovávání v archivu). Vzhledem k tomu, že výstupy projektu InterPARES představují de facto jedinou komplexní konceptualizaci autenticity určitého typu dokumentu digitálního dědictví vůbec, budou tyto výstupy představeny podrobněji a rozděleny do tematických podsekcí pro větší přehlednost.

#### **5.2.2.1 Model pro elektronickou archiválii**

Rovněž projekt InterPARES založil svůj model dokumentu na konceptech klasické diplomatiky. Výsledkem je však odlišný model, který je specificky modelem elektronické archiválie. Model se skládá ze tří konstitutivních součástí (*constituent parts*) (73, s. 5):

- dokumentová forma (*documentary form*)
- anotace (*annotations*)
- kontext (*kontext*)

Dokumentová forma jsou „pravidla reprezentace, na základě kterých je komunikován obsah záznamu, jeho bezprostřední správní a dokumentový kontext a jeho pravomoc (*authority*)“ (73, s. 5). Dokumentová forma zahrnuje vnitřní znaky (*intrinsic elements*) a vnější znaky (*extrinsic elements*). Vnitřní znaky jsou specifické pro archivářský kontext, vnější znaky lze aplikovat i na jiné typy dokumentů. Vnější znaky tvoří: a) celkové prezentační charakteristiky (např. text, obraz, zvuk); b) specifické prezentační charakteristiky (např. rozložení textu, barva, prostorové rozlišení); c) elektronické podpisy a elektronická razítka (např. digitální podpisy); d) digitální časové známky; e) speciální prvky (např. digitální vodoznak) (73, s. 6). Anotace jsou dodatky, které byly přidány k archivnímu dokumentu až poté, co již byl vytvořen (76).

Kontexty záznamu se dělí na: a) dokumentový kontext (archivní sbírky, do kterých dokument patří, a vnitřní struktura záznamu); b) procedurální kontext (proces, v jehož průběhu byl záznam vytvořen); c) technologický kontext (charakteristické vlastnosti technických komponent počítačového systému, ve kterém je dokument vytvářen); d) provenienční kontext (původce dokumentu, jeho mandát, organizace a funkce); e) právně-správní kontext (právní a organizační systém, jehož je původce součástí) (77, s. 3).

Otázka média (*medium*) byla pro členy výzkumného týmu sporná. Pracovní skupina pro autenticitu označila médium za čtvrtou konstitutivní součást elektronické archiválie (73, s. 5), zatímco druhá pracovní skupina (věnující se uchovávání) dospěla k závěru, že „médium by mělo být považováno za součást technologického kontextu dokumentu, a nikoliv za nezbytnou součást samotného archivního dokumentu.“ (73, s. 7). Argumentace této skupiny byla založena na těchto skutečnostech: a) médium a sdělení již nejsou neodlučně spojené (potud shoda s předchozím projektem); b) to, co je zapsáno na médiu, není dokument jako takový, ale bitový tok; c) výběr média je obvykle arbitrární a nemá žádný zvláštní význam (73, s. 6). Médium je podle druhé pracovní skupiny týmu sice nutné k tomu, aby archivní dokument (jako bitový tok) mohl vůbec přetrvávat, ale není dostatečné k tomu, aby bylo možno tento dokument uchovat jako dokument.

#### **5.2.2.2 Uchovávání elektronických archiválií**

Z hlediska otázky uchovávání archiválií dospěl výzkumný tým InterPARES k závěru, který je užitečné vzhledem k jeho obecné povaze (lze jej aplikovat i na jiné typy dokumentů) uvést celý: „uchovávat elektronickou archiválii není možné: lze pouze uchovávat schopnost tuto archiválii reprodukovat (*reproduce*)... Mezi digitální reprezentací dokumentu v úložišti a formou, ve které je prezentován k užití, nutně existuje zásadní rozdíl. Vždy je nutné využít

nějaký software k převedení uložených digitálních bitů do dokumentové formy archiválie. To nevyhnutelně vytváří rizika, že bez ohledu na to, jak dobře byla digitální data zabezpečena na úložišti, může být archiválie nevhodně pozměněna v procesu načítání uložených bitů a prezentování uživatelům... proces uchovávání elektronické archiválie přesahuje rámec zabezpečeného uložení... začíná prvotním aktem uložení dokumentu a pokračuje jeho reprodukováním.“ (78, s. 5). Tento závěr znamená, jak sami výzkumníci uvádějí, paradigmatický posun v přístupu k uchovávání. Jak je na první pohled zřejmé, jde o zcela opačný závěr, než který byl výstupem předchozího projektu.

Projektový tým dále vymezil elektronickou archiválii jako celek jedné nebo více digitálních komponent, přičemž každá komponenta je definována tím, čím se od jiné odlišuje technologicky, způsobem uložení bitů nebo metodami (softwarem), které musejí být použity k reprodukování úplného dokumentu (78, s. 5-6). Reprodukování elektronické archiválie technologií zahrnuje její rekonstrukci (opětovné složení jejích digitálních komponent) a prezentování v adekvátní formě (78, s. 6).

Proces digitální archivace uchovávání podle týmu zahrnuje dva primární výstupy: reprodukované (*reproduced*) archiválie a reprodukovatelné (*reproducible*) archiválie. První výstup znamená, že elektronický systém pro správu dokumentů přímo zajišťuje (pomocí procesů rekonstrukce a prezentace) reprodukci archiválií svým uživatelům v rámci organizace. Druhý výstup plní požadavek zpřístupnit elektronické archiválie mimo tento systém (např. v internetovém prohlížeči uživatele) a znamená export reprodukovatelné archiválie (tj. poskytnutí všech jejích digitálních komponent spolu s instrukcemi, jakým způsobem z nich rekonstruovat a prezentovat archiválii). V tomto popisu je nutno rozpoznat specifický způsob uložení a práce s archiváliemi v organizacích, kdy jsou tyto dokumenty spravovány v systémech pro spisovou službu nebo podobných interních řešeních, které obvykle nemají webové rozhraní.

V procesu digitální archivace nutně vznikají kopie dokumentů (výstupy replikace nebo konverze), které výzkumný tým popisuje následovně: „Kopie může být vytvořena z originálu, nebo z kopie, a to buď originálu nebo jiné kopie“ (79, s. 4). Originál (*original record*) je „první, úplný dokument, který je schopný dosáhnout svého účelu“, tj. v rámci činnosti původce (79, s. 4), v terminologickém slovníku se uvádí definice „první úplný a účinný dokument“ (76, s. 5). Elektronické archiválie předávané původcem do archivu jsou tedy vždy z principu již jen kopie dokumentů. Podle týmu „nejspolehlivější kopie je kopie vytvořená z originálu, která je identická s originálem, ačkoliv byla vytvořena později. Imitativní kopie

(*imitative copy*) je kopie, která převádí obsah i formu dokumentu, ale takovým způsobem, že lze vždy poznat, co je kopie a co originál. Jednoduchá kopie (*simply copy*) je kopie, která převádí pouze obsah originálu.“ (79, s. 4).

Jistým problémem je, že výstupy projektu obsahují výraz „reproduction“, který je v nich užíván ve dvou odlišných významech. V prvním významu jako „proces vytváření kopií“ (79, s. 4), v druhém významu (uvedeným přímo v terminologickém slovníku projektu) jako „sestavení dat ve srozumitelné formě pro komunikaci“ (76). V překladu jsem tuto dvojznačnost odstranil tím, že překlad „reprodukce“ používám v celé práci jednotně v druhém významu.

### 5.2.2.3 Koncept autenticity elektronických archiválií

Ve svém terminologickém slovníku projekt definuje autentický archivní dokument jako „archivní dokument, který je tím, za co se vydává (*what it purports to be*), a který nebyl zfalšován (*tampering*) nebo porušen (*corruption*)“ (76). Ačkoliv tato definice přibližně odpovídá definici předcházejícího projektu, konceptualizace autenticity v projektu InterPARES se liší vyšší mírou komplexity, explicitním připuštěním možnosti určitých změn dokumentu bez narušení autenticity a problematizací tradičního principu, podle kterého skutečnost, že se původce spoléhá na své procesy vytváření kopií dokumentů, činí tyto kopie dokumentů autentickými.

Výzkumníci projektu uvádějí, že digitální technologie přinášejí podstatně větší riziko (záměrného i neúmyslného) ohrožení autenticity: „autenticita elektronických dokumentů je ohrožena, kdykoliv jsou přenášeny prostorem (tj. odesílány adresátovi nebo mezi systémy a aplikacemi) nebo časem (tj. během uložení v úložišti nebo aktualizace či náhrady hardwaru nebo softwaru užívaného k jejich ukládání, zpracování nebo komunikaci.“ (79, s. 2). To se vztahuje na fázi správy aktivních dokumentů v rámci činnosti původce i na přenos neaktivních dokumentů (jako archiválií) do archivu a jejich následné dlouhodobé uchovávání. Projekt stanovuje požadavky na dvě dílčí fáze činnosti archivu:

1. přijetí archiválií od původce (sada požadavků A), kdy je úkolem archivu posoudit jejich autenticitu;
2. dlouhodobé uchovávání těchto archiválií v archivu (sada požadavků B), kdy archiv musí zajistit udržování autenticity přijatých archiválií v průběhu času.

V první dílčí fázi musí archiv posouzení autenticity podložit „dokladem (*evidence*), že dokument je tím, za co se vydává a že nebyl pozměněn (*modified*) nebo porušen

v podstatných ohledech (*in essential respects*).“ (79, s. 1). Skutečnost, že dokumenty používal ke své činnosti původce, tedy podle tohoto výzkumu již není dostačující záruka jejich autenticity. Posouzení autenticity elektronického dokumentu znamená „zjistit (*establish*) jeho identitu a prokázat (*demonstrate*) jeho integritu“ (79, s. 1).

Identita dokumentu jsou „atributy dokumentu, které jej jedinečně charakterizují a odlišují od jiných dokumentů“ (79, s. 1). Integrita se týká úplnosti (*wholeness*) a neporušenosti (*soundness*) dokumentu: „dokument má integritu, pokud je úplný a neporušený ve všech podstatných aspektech (*essential aspects*).“ (79, s. 2). Podle týmu to znamená splnění podmínky, 1) aby „nebylo pozměněno sdělení, které má dokument přenášet s cílem dosáhnout určitého účelu“, a 2) aby „zůstaly zachovány vyjádření obsahu a všechny požadované anotace a znaky dokumentové formy“ (79, s. 2). Jako příklad je uváděna autentická kopie e-mailu, která sice při zobrazení vypadá jinak, než jak ji viděl původní příjemce, ale která zachovává obsah původního e-mailu a informace o odesílateli, datu odeslání apod. Znamená to, že je možno zachovat integritu (a tím autenticitu) při změně bitů digitálních komponent dokumentu, pokud byly splněny výše uvedené podmínky.

Atributy identity mohou být explicitně vyjádřeny v prvcích dokumentu, v souvisejících metadatech nebo mohou být implicitní (tj. vyvozené z různých kontextů dokumentu). Integritu lze prokázat doklady získanými ze samotného dokumentu, ze souvisejících metadat nebo z jednoho či více kontextů dokumentu (79, s. 2).

#### **5.2.2.4 Požadavky na posuzování autenticity přebíraných elektronických archiválií**

Při posuzování autenticity elektronických archiválií musí archiv vyhodnotit způsob, jakým původce spravoval své dokumenty od okamžiku jejich vytvoření. Jednou ze dvou metod, kterou podle výzkumníků projektu InterPARES může archiv užít pro toto posuzování, je tzv. *presumpce autenticity* (*presumption of authenticity*). Presumpce autenticity je „vyvozování ze skutečností známých o způsobu vytváření a udržování dokumentů.“ (79, s. 3).

Pro posuzování autenticity touto metodou byl vytvořen následující seznam požadavků („sada A“) (79, s. 5-6):

<b>POŽADAVKY PRO POSOUZENÍ AUTENTICITY DOKUMENTU („SADA A“)</b>			
<b>A.1</b>	<b>Vyjádření atributů a vazeb dokumentu</b>		
	<b>A.1.a</b>	<b>Identita dokumentu</b>	
		A.1.a.i	Jména osob spolupůsobících při vytváření dokumentu
		A.1.a.ii	Název úkonu nebo záležitosti, k níž se dokument vztahuje
		A.1.a.iii	Datace vytvoření a přenosu dokumentu, např. datum vytvoření v systému; datum přijetí adresátem; datum uložení do sbírky; datum přenosu (tj. opuštění prostoru, ve kterém byl dokument vytvořen)
		A.1.a.iv	Vyjádření archivní vazby (např. klasifikační kód)
		A.1.a.v	Údaje o přílohách
	<b>A.1.b</b>	<b>Integrita záznamu</b>	
		A.1.b.i	Název subjektu, který je formálně způsobilý k zajištění úkonu nebo záležitosti, ke které se dokument vztahuje
		A.1.b.ii	Název subjektu, který má hlavní odpovědnost za vytvoření směrodatné kopie dokumentu
		A.1.b.iii	Údaje o anotacích přidaných k dokumentu
		A.1.b.iv	Údaje o technických modifikacích dokumentu

**Tabulka 1 - Požadavky pro posouzení autenticity: část 1/2 (InterPARES)**



<b>POŽADAVKY PRO POSOUZENÍ AUTENTICITY DOKUMENTU („SADA A“)</b>	
A.2	<b>Přístupová práva</b>
	Původce určil a účinně zavedl přístupová práva týkající se vytváření, modifikací, anotací, přesunu a skartace dokumentů.
A.3	<b>Ochranné procedury: ztráta a poškození dokumentu</b>
	Původce stanovil a účinně zavedl procedury pro zamezení, odhalení nebo opravu ztráty nebo poškození dokumentů
A.4	<b>Ochranné procedury: datové nosiče a technologie</b>
	Původce stanovil a úspěšně zavedl procedury pro zabezpečení identity a integrity dokumentů před degradací nosičů a technologickými změnami.
A.5	<b>Stanovení dokumentových forem</b>
	Původce stanovil dokumentové formy archivních dokumentů pro každou proceduru v souladu s požadavky právního řádu, nebo požadavky původce.
A.6	<b>Autentifikace dokumentu</b>
	Původce stanovil konkrétní pravidla vztahující se k tomu, jaké dokumenty musejí být autentifikovány, kým a jakými prostředky (pokud je vyžadována právním řádem nebo potřebami organizace).
A.7	<b>Určení směrodatné kopie dokumentu</b>
	V případě, že existuje více kopií téhož dokumentu, původce stanovil procedury, pomocí nichž lze určit, jaká kopie dokumentu je směrodatná.
A.8	<b>Odstranění a předání relevantní dokumentace</b>
	Původce stanovil a účinně zavedl procedury, které určují, jaká dokumentace musí být odstraněna a převedena do archivu spolu s dokumenty.

**Tabulka 2 - Požadavky pro posouzení autenticity: část 2/2 (InterPARES)**

Požadavky „sady A“ jsou kumulativní povahy (čím více požadavků a do čím vyšší míry je splněno, tím silnější je presumpce autenticity).

Požadavky sekce „A.1“ představují podle týmu základní nezbytné informace pro zjištění identity dokumentu a představují základ pro prokázání jeho integrity. Hodnoty těchto atributů musejí být vyjádřeny explicitně a neodlučně spojeny s každým dokumentem po celou dobu jeho životního cyklu. Jedním z hlavních rizik je špatně nastavený export těchto atributů ze systému původce. Údaje o technických modifikacích dokumentu (A.1.b.iv) zahrnují jakékoliv změny digitálních komponent (formátu, metod reprodukce apod.), které mohou vyvolávat otázky, zdali je reprodukován dokument stejný, jako by byl, pokud by k modifikaci nedošlo (79, s. 6). Tyto údaje mohou také odkazovat na dodatečnou externí dokumentaci, která podrobněji vysvětluje povahu těchto modifikací. Po určitém zobecnění lze v attributech sekce „A.1“ spatřit částečnou paralelu s praxí bibliografického popisu.

Dalších sedm požadavků (A.2 až A.8) se týká podpory presumpce integrity a popisuje procedurální řízení životního cyklu dokumentu. Autentifikací (*authentication*) dokumentu se zde myslí prohlášení kopie dokumentu za autentickou v určitý okamžik času právníkou osobou s pravomocí takové prohlášení vykonat (tedy vidimace v českém právním kontextu) (79, s. 9). Autentifikaci kopií tým odlišuje od validace (*validation*) procesu kopírování digitálních komponent dokumentu. Tato validace se uskutečňuje pokaždé, když jsou dokumenty replikovány na jiný datový nosič nebo konvertovány z jedné technologie do druhé (čímž se v kontextu projektu myslí konverze všech dokumentů z jednoho systému pro správu dokumentů, např. spisové služby, do jiného).

#### **5.2.2.5 Požadavky na udržování autentických elektronických archiválií**

Archiv podle týmu musí: zavést a dodržovat postupy zajišťující, aby autenticita dokumentů nebyla ohrožena procesem vytváření kopií; vytvářet a udržovat dokumentaci vztahující se ke způsobu, kterým udržoval záznamy v průběhu času, a ke způsobu, kterým vytvářel kopie. Pro vlastní dlouhodobé uchovávání elektronických dokumentů vytvořil projektový tým druhou sadu požadavků („sada B“), které se vztahují k udržování autenticity archiválií převzatých od původce. Archiv musí plnit všechny následující požadavky (79, s. 7-8).

<b>POŽADAVKY PRO UDRŽOVÁNÍ AUTENTICKÝCH ARCHIVÁLIÍ („SADA B“)</b>		
B.1	<b>Řízení převodu, správy a procesu vytváření kopií</b>	
	Procedury a systémy pro převod dokumentů do archivu, správu dokumentů v archivu a procesy vytváření kopií v archivu musejí mít náležité a účinné kontrolní mechanismy, které zajišťují:	
	B.1.a	udržení nepřerušené správy dokumentů
	B.1.b	zavedení a monitorování zabezpečovacích a kontrolních procedur
	B.1.c	zachování obsahu dokumentu, všech vyžadovaných anotací a znaků dokumentové formy po skončení procesu vytváření kopií
B.2	<b>Dokumentace procesu kopírování a jeho důsledků</b>	
	Proces vytváření kopií musí být zdokumentován, přičemž tato dokumentace musí obsahovat:	
	B.2.a	Datum vytvoření kopie dokumentu a jméno odpovědné osoby
	B.2.b	Vztah mezi dokumenty získanými od původce a kopiemi těchto dokumentů vytvořenými archivem
	B.2.c	Důsledky procesu vytváření kopií na formu, obsah, zpřístupnitelnost a užití dokumentů
	B.2.d	Případné informace o tom, že kopie dokumentu nepřenáší v úplnosti a věrně všechny znaky vyjadřující identitu a integritu původního dokumentu; tato dokumentace by měla být pro uživatele snadno dostupná
B.3	<b>Archivní popis</b>	
	Archivní popis sbírek dokumentů musí zahrnovat popis kontextů dokumentů a informace o změnách, kterými prošly od doby svého vytvoření	

**Tabulka 3 - Požadavky pro udržování autenticity (InterPARES)**

Ve většině požadavků se opakují požadavky „sady A“. Požadavek „B.1.a“ zahrnuje řízení převodu dokumentů od původce a ověření, zda byly do archivu spolu s dokumenty převedeny i všechny atributy identity a integrity („A.1“) a veškerá relevantní dokumentace („A.8“). Požadavek „B.1.b“ zahrnuje zejména určení přístupových práv („A.2“) a vykonávání

ochranných procedur („A.3“ a „A.4“), které by měly být efektivně zavedeny a pravidelně monitorovány. Požadavek „B.1.c“ znamená stanovení, zavedení a monitorování procedur vytváření kopií, které dokáží zajistit, aby se obsah záznamu během vytváření kopií nezměnil.

K požadavku „B.2“ se uvádí, že dokumentace je „nezbytným prostředkem, jak doložit, že proces vytváření kopií je transparentní (tj. bez předstírání nebo klamu).“ (79, s. 11). Je důležitá jak z hlediska archivu (doklad důvěryhodnosti jakožto správce záznamů), tak pro uživatele: „jelikož historie vytváření kopií je nezbytnou součástí historie dokumentu. Dokumentace procesu a jeho důsledků poskytuje uživatelům dokumentů kritický nástroj pro jejich posouzení a interpretování.“ (79, s. 11).

Tým spatřuje stěžejní význam v užití tradičního konceptu archivního popisu (tj. sbírek archiválií) („B.3“) pro kontext elektronických archiválií: „archivní popis shrnuje historii všech vytváření kopií, čímž snižuje potřebu uchovávat dokumentaci všech jednotlivých aktů vytváření kopií. V tomto ohledu popis ustanovuje hromadné potvrzení autenticity záznamů a jejich vztahů v kontextu sbírek, do kterých patří.... Důležitost tohoto kolektivního osvědčení je ta, že autentifikuje a zachovává vztahy mezi dokumenty téže sbírky.“ (79, s. 11).

### 5.3 Autenticita v normě ISO 14721

V prvním vydání normy ISO 14721 (2003) není pojem autenticity (ve smyslu autenticity dokument) definován. První vydání normy pracuje pouze se substantivem „authentication“, a to ve dvou významech. Prvním je autentizace uživatele, tj. běžný význam tohoto pojmu v oblasti řízení bezpečnosti informací (v tomto případě norma užívá též sloveso „authenticate“). Druhý význam je svázaný s informacemi o neporušenosti (*fixity information*), které jsou v prvním vydání normy definovány jako „informace, které dokumentují kontrolní mechanismy (*authentication mechanisms*) a poskytují kontrolní klíče (*authentication keys*) zajišťující, aby objekt s informačním obsahem nebyl nezdokumentovaným způsobem změněn“ (12, s. 1-10). Jako příklad je uveden kód CRC; tímto druhým významem je tedy hašovací funkce. Výraz „authenticity“ je v textu normy užit pouze jednou: v úvodním přehledu historie uchovávání se uvádí, že tradiční archivy spravují své archiválie tak, „aby byl uchován jejich informační obsah a autenticita.“ (12, s. 2-1).<sup>60</sup> Pátá povinnost archivu OAIS je zde vymezena jako: „dodržovat zdokumentovaná pravidla a postupy, které zajistí, že

---

<sup>60</sup> Dále je ještě užit v tabulce příkladů implementace archivačních informací, ve spojení „indikátor autenticity“ (*authenticity indicator*) právě v souvislosti s informacemi o neporušenosti (12, s. 2-1).

informace budou chráněny před všemi možnými nepředvídatelnými událostmi, a které umožní šíření informací v podobě autentifikovaných kopií (*authenticated copies*) originálu nebo v podobě, ze které lze originál zpětně určit“ (12, s. 3-1). Norma však nedefinuje, co se „autentifikovanými kopiemi“ myslí, ani jak má být tohoto cíle dosaženo. Z textu je zřejmé, že nemůže jít o autentizaci uživatele. Souvislost s informacemi o neporušenosti je také nepravděpodobná. Podle všeho tato část odkazuje na právně platné ověření digitálních kopií, o kterém bylo pojednáno v popisu projektu InterPARES. Je pravděpodobné, že byl tento koncept převzat z tohoto projektu.

Sám David Giaretta (jeden z hlavních autorů normy ISO 14721) ve své rozsáhlé publikaci o modelu OAIS přiznává, že absence tematizace autenticity v prvním vydání normy byl její hlavní nedostatek (80, s. 205). Jeho kniha také naznačuje, že nové vydání normy ISO 14721 reflektuje výstupy projektu InterPARES.

V druhém vydání normy se výraz „authentication“ (příp. „authenticate“) uvádí již pouze v souvislosti s autentizací uživatele. Definice informací o neporušenosti byla změněna tak, aby tento výraz již neužívala pro popis hašovací funkce.<sup>61</sup> Především je zde však autenticita definována a zavedena do celkové logiky normy. Definice je: „Míra, do níž osoba (nebo systém) pokládá objekt za ten, za který je vydáván (*The degree to which a person (or system) regards an object as what it is purported to be.*); autenticita se posuzuje podle důkazů (*evidence*)“ (10, s. 20). Úkol zachovat autenticitu je zanesen i do samotného pojmu „dlouhodobé uchovávání“ (*long term preservation*), což je nově: „dlouhodobé udržování informací v podobě, která je cílové komunitě srozumitelná sama o sobě, spolu s doklady (*evidence*) o jejich autenticitě“ (10, s. 24). Druhá část popisu páté povinnosti archivu OAIS původní normy je v novém vydání přesunuta do popisu šesté povinnosti a upravena tak, že nová definice této povinnosti je již srozumitelná: „Zpřístupňovat uchovávané informace cílové komunitě a umožňovat šíření informací v podobě kopií původně dodaných datových objektů nebo v takové podobě, aby bylo možné zpětně dohledat (*as traceable to*), ke kterým původně dodaným datovým objektům se vztahují, a to společně s doklady o jejich autenticitě.“ (10, s. 39). Nevysvětlený výraz „autentifikované kopie“ prvního vydání byl nahrazen uvedením nutnost uchovávat doklady o autenticitě. Archiv má tedy uchovávat i zpřístupňovat informace spolu s doklady o jejich autenticitě.

---

<sup>61</sup> Zůstal pouze uvedený příklad indikátor autenticity, který zřejmě nebyl při redakci textu odstraněn (10, s. 75)

Norma k těmto dokladům uvádí: „Doklady o autenticitě poskytuje producent jako součást dodávaných archivačních informací. Archiv pak tyto doklady v průběhu času udržuje, aktualizuje nebo doplňuje. Jako součást provenienčních informací může producent dodat popis vlastností informací pro ty vlastnosti informací, které by měly být v průběhu času udržovány (nebo tento popis může vytvořit přímo archiv). Producent samozřejmě může dodat i popisy vlastností informací těch vlastností informací, které v průběhu času není třeba udržovat.“ (10, s. 50) Vlastnostmi informací norma specificky rozumí vlastnosti informačního obsahu (nikoliv ostatních typů informací, jako jsou např. archivační informace).

V případě, že se archiv rozhodne provést nevratnou transformaci objektu CDO (například formátovou konverzi z nekomprimovaného formátu TIFF do JP2 v matematicky ztrátové kompresi), norma doporučuje stanovit transformační vlastnost informací (*transformational information property*), kterou definuje následovně: „vlastnost informací, u které je uchování její hodnoty pokládáno za nutné, ale ne dostačující, aby bylo možné ověřit, zda při jakékoliv nevratné transformaci byl zachován informační obsah“ (10, s. 27). Taková vlastnost informace je podle normy závislá na konkrétních interpretačních informacích (10, s. 106-107) a její zachování může „podstatným způsobem dokládat autenticitu“ (10, s. 24). Transformační vlastnost informací je tak podle normy specifickým typem vlastnosti informace, který hraje roli při návrhu nevratné transformace a posuzování jejího výsledku. Norma upozorňuje, že tato vlastnost je závislá na konkrétní reprezentaci a uvádí následující příklad transformace digitální knihy (10, s. 106-107). Digitální kniha má (po zobrazení softwarem) podobu stránek s okraji, názvem, názvy kapitol, odstavci a řádky, který jsou složeny ze slov a interpunkce. Pokud jsou za transformační vlastnosti informací stanoveny „odstavce a znaky vyjadřující slova a interpunkci“, pak je možno ostatní vlastnosti změnit, a přitom bude zachována požadovaná úroveň uchovávání. Tento příklad také dobře ilustruje uvedenou definici informačního obsahu,<sup>62</sup> která zahrnuje variantu, že předmětem uchovávání může být pouze část původního informačního obsahu (tj. to, co je výsledkem převodu pouze některých vlastností).

Koncept nutnosti udržování dokladů o autenticitě při správě dokumentů v digitálním archivu norma ISO 14721 nepochybně převzala z výzkumu InterPARES. Norma však nepopisuje žádnou konkrétní sadu požadavků na udržování autenticity: „Archiv může na základě praxe a doporučení dané komunity (včetně osvědčených postupů, směrnic, standardů a právních náležitostí) vyhodnotit autenticitu svých jednotek. Vědecké archivy

---

<sup>62</sup> Viz oddíl 2.2.2.1

mohou mít například méně přísná hodnoticí kritéria než státní archivy. Koncový uživatel však může vyhodnotit autenticitu sám, a to nejprve podle toho, jak je doložena v archivačních informacích“ (10, s. 50). Norma pouze uvádí, že součástí dokladů o autenticitě tvoří provenienční informace (které mohou zahrnovat popis vlastností informací, které je nutno zachovat) (10, s. 25).

Pojem „transformační vlastnost informací“ je další novinkou druhého vydání normy. Koncept provenienčních informací byl obsažen již v prvním vydání, ale bez souvislosti s autenticitou. Specifikem normy ISO 14721 je důraz na to, že autenticita je otázkou míry.

Tento aspekt dále vhodně rozvádí navazující norma ISO 16363. Jako komentář k požadavku zpřístupňování informací spolu s doklady o jejich autenticitě uvádí: „Autenticita není nějaký absolutní koncept, ale je otázkou míry a posuzuje se podle dokladů (*evidence*). Z tohoto důvodu má při posuzování tohoto požadavku klíčový význam patřičná dokumentace (*evidence*). Tento požadavek vyžaduje, aby opatření vztahující se k příjmu, uchovávání a převodu nevedly ke ztrátě těch informací, které by mohly posloužit jako prověřitelný záznam autenticity mezi původním přijatým objektem a konečným poskytovaným objektem.“ (11, s. 62). Archiv by měl mít podle této normy dokumentaci procesů vytváření balíčků DIP z balíčků AIP; takováto dokumentace je klíčová pro zjišťování „ověřitelného řetězce autenticity (*auditable chain of authenticity*) vedoucího od balíčku AIP k zpřístupňovaným digitálním objektům“. Z textu normy ISO 14721 lze vyvodit, že v praxi budou hlavní podobou dokladů o autenticitě metadata zaznamenávající archivační informace. V souladu s požadavky normy ISO 16363 pak musejí být tyto doklady také v podobě dokumentace (důležitý je zejména dokument popisující pravidla pro uchovávání).

## 5.4 Autenticita ve standardu PREMIS

Standard PREMIS je klíčový standard digitální archivace. Sám jako jednu ze svých funkcí uvádí podporu autenticity digitálního objektu v kontextu archivačního procesu (52, s. 1). Tento standard je praktickou aplikací konceptu provenienčních informací modelu OAIS, tj. specifikací způsobů, jakým je lze zaznamenat.

PREMIS uvádí pojem autenticita v souvislosti s integritou a neporušeností (*fixity*). Neporušeností se zde rozumí „vlastnost digitálního objektu, která udává, že objekt nebyl změněn mezi dvěma body v čase.“ (52, s. 270). Pro záznam této vlastnosti jsou vyčleněny sémantické jednotky pro název algoritmu digitálního otisku, hodnotu digitálního otisku

a název činitele, který jej vytvořil. Kontrola neporušenosti by podle PREMIS měla být zapsána jako událost s výsledkem. Lze ji zapsat pouze na úrovních souboru a bitového toku.

Integrita v tomto standardu definována není. Ověření integrity je událost, která se vztahuje k souboru a reprezentaci. Pro ověření integrity souboru jsou klíčové mechanismy formátová identifikace a validace. Pro ověření integrity reprezentace jsou nutné speciální programy, které rozumí struktuře reprezentace, přičemž k této operaci může být potřeba využít strukturální metadata. Standard tím poukazuje zejména na potřebu speciálních validátorů, které musejí být vytvořeny nebo nastaveny tak, aby byly schopny testovat strukturu konkrétních typů dokumentů dané instituce (např. právě digitalizátů knihy vytvářených dle daného strukturálního modelu reprezentace).

Autenticita je pak podle PREMIS „vlastnost, že digitální objekt je tím, za co se vydává; tj. že integritu zdroje i obsahu digitálního objektu lze ověřit.“ (52, s. 267). Klíčovým prvkem autenticity je digitální provenience (52, s. 15), chápána jednoduše jako „historie objektu“ (52, s. 2). Standard dále definuje autentifikaci (*authentication*) jako „prokázání autenticity“ (52, s. 259). Pro autentifikaci jsou podle standardu nutné jisté technické a procedurální doklady, ačkoliv je autentifikace v konečné míře otázkou úsudku. Jako příklady autentifikace uvádí (52, s. 259):

- udržování podrobné dokumentace digitální provenience;
- uchovávání kopie objektu, která je identická s původně dodaným objektem;
- zachování signifikantních vlastností;
- využití digitální podpisů.

Pro digitální provenienci je klíčovým prvkem, který musí být zaznamenáván, vztah odvození (*derivation*), který je vždy vztahem mezi předcházejícím a novým objektem, přičemž odvození je definováno jako „výsledek replikace nebo transformace objektu“ (52, s. 19). Z odvození jsou za rozhodující považovány ty, které pozměňují digitální objekt; jejich záznam je tedy pro zachování autenticity nezbytností (52, s. 19).

Jak bylo uvedeno v přehledu PREMIS, logika tohoto standardu předepisuje jakoukoliv změnu digitálního objektu zaznamenat jako vztah odvození nového objektu z předchozího objektu (tzn. vlastností objektu nemůže být jeho verze). Původní objekt může být smazán, a pak po něm zůstane jen „metadatová stopa“ v digitální provenienci.



Následující ukázka je metadatový záznam objektu B, který vznikl formátovou konverzí z objektu A:

```
relationshipType = "derivation"
relationshipSubType = "has source"
relatedObjectIdentifier
relatedObjectIdentifierType = "local"
relatedObjectIdentifierValue = "A"
relatedEventIdentifier
relatedEventIdentifierType = "local"
relatedEventIdentifierValue = "1"
```

V případě odvození musí být událost spojena s objektem pomocí identifikátoru související události (*relatedEventIdentifier*). Jiné typy událostí, které nevytvářejí jiný objekt (např. kontrola neporušenosti), jsou s objektem propojeny pomocí identifikátoru přidružené události (*linkingEventIdentifier*), příkladem je kontrola neporušenosti. Dále je možno zapsat více událostí na témže objektu a určit jejich posloupnost, a to buď přímo (užitím pořadového čísla), nebo nepřímo (užitím datace události). Uchovávání identické kopie znamená identitu na úrovni bitů s objektem, který do archivu dodal vkladatel.

Signifikantní vlastnosti (*significant properties*) jsou „vlastnosti jednotlivého objektu, u nichž je subjektivně stanoveno, že je důležité je udržovat napříč archivačními opatřeními... i když mají objekty stejné technické vlastnosti, mohou se lišit vlastnostmi, které by měly být zachovány pro budoucí prezentaci nebo užití.“ (52, s. 50). Tyto vlastnosti odpovídají transformačním vlastnostem modelu OAIS.

Signifikantní vlastnosti se vztahují na všechny úrovně objektu (intelektuální entita, reprezentace, soubor, bitový tok) a jejich elementy jsou:

- *significantPropertiesType*
- *significantPropertiesValue*
- *significantPropertiesExtension*

Standard upozorňuje, že v některých případech musí hodnoty dodat vkladatel nebo kurátor archivu. Některé vlastnosti jsou měřitelné, jiné subjektivní. Uvedení těchto vlastností

znamená, že archiv má v úmyslu tyto vlastnosti zachovat, a jiné nikoliv. Stanovení signifikantních vlastností může být na různém stupni podrobnosti (viz obsah vs. počet stran). Standard přiznává obtížnost definování těchto vlastností a potřebu dalšího výzkumu.

V praxi se tyto elementy zřídka užívají (ostatně nejsou povinné) – je totiž obtížné je nejen identifikovat (tj. určit, co musí být zachováno), ale také definovat (otázka, jak přesně takovou vlastnost vyjádřit). Třetí element této sekce umožňuje vnořit externí schémata pro popis signifikantních vlastností. Pro digitalizáty knih však žádné takové schéma neexistuje.

Digitální podpis (*digital signature*) je definován jako „hodnota, která je vypočítána kryptografickým algoritmem a připojena k datům takovým způsobem, aby každý příjemce dat mohl tento podpis užít k ověření původu dat a integrity.“ (52, s. 269). Tento podpis lze podle standardu použít v archivu následujícími způsoby: při příjmu dat do archivu (potvrzení identity vkladatele); při šíření dat z archivu (potvrzení identity archivu); při uložení (uchovávání podepsaných objektů pro potvrzení původu a integrity dat). V prvních dvou případech není důvod digitální podpis po skončení validace uchovávat.

## 5.5 Autenticita v knihovnictví

Ve zdrojích z oblasti knihovnictví nenalezneme žádnou konceptualizaci autenticity digitalizovaných knih, pouze její tematizaci jako součást širších okruhů a kontextů. Jimi jsou: archivace (obecné otázky autenticity z hlediska uchovávání), digitalizace (otázky autenticity spojené s produkcí digitalizátů) a zpřístupňování (otázky autenticity dokumentu jakožto zpřístupňovaného dokumentu).

### 5.5.1 Autenticita a uchovávání

Priscila Caplanová ve svém článku „The preservation of digital materials“ shrnula přehled problematiky digitální archivace knihovních dokumentů do sedmi cílů (8, s. 7-9):

Stupeň	Cíl	Popis cíle	Rizika	Ochranná opatření
1	Dostupnost ( <i>availability</i> )	Kontrola nad digitálním objektem	Nemožnost provádět některá ochranná opatření (instituce objekt nevlastní apod.)	Výběr ( <i>selection</i> ) / Převzetí kontroly nad digitálním objektem ( <i>capture</i> )
2	Identita ( <i>identity</i> )	Popis digitálního objektu	Ztráta sebe-popisnosti digitálního objektu	Spojení popisu s objektem samým, trvalé identifikátory
3	Srozumitelnost ( <i>understandability</i> )	Informace o širším kontextu digitálního objektu	Nedostatek informací o kontextu	Dokumentace kontextu
4	Neporušenost ( <i>fixity</i> )	Zajištění neměnnosti digitálního objektu	Neautorizované změny (záměrné, neúmyslné)	Zabezpečené uložení ( <i>secure storage</i> )
5	Životnost ( <i>viability</i> )	Zajištění možnosti načíst digitální objekt z datového nosiče	Zastaralost nebo degradace datového nosiče	Správa datových nosičů ( <i>media management</i> )
6	Reprodukovatelnost ( <i>renderability</i> )	Zpřístupnění digitálního objektu (softwarem)	Formátová zastaralost	Formátové strategie ( <i>format strategies</i> )
7	Autenticita			

Tabulka 4 - Sedm cílů digitální archivace (Caplanová)

Sedmým (nejvyšším) cílem archivace je autenticita, o které pojednáme blíže. Autenticity podle Caplanové „znamená, že lze ověřit integritu zdroje i obsahu objektu...

Musejí existovat pravidla a postupy, které zabezpečí integritu dat (tj. že objekty nejsou zničeny nebo neautorizovaně pozměněny) a které zajistí dokumentaci řetězce správy a všech autorizovaných změn. Historie událostí vztahujících se k digitálnímu objektu je známá jako jeho „digitální provenience“ a je rozhodující součástí archivačních metadat.“ (8, s. 8).

Z textu není zřejmé, co se myslí zdrojem objektu (datový objekt?). Stupně cílů by měly naznačovat, že splnění vyššího cíle předpokládá splnění všech nižších. Zajištění reprodukovatelnosti je však podle Caplanové „srdcem procesu digitální archivace“ a objekt podle ní může být autentický, i když jej nelze reprodukovat (8, s. 9). Podle této autorčiny logiky by tedy nejvyšším cílem měla být spíše reprodukovatelnost, nikoliv autenticita.

Marilyn Deeganová a Simon Tanner se v úvodní stati o digitální archivaci věnují srovnání procesů uchovávání fyzických a digitálních dokumentů (81). Autenticitu obou typů dokumentů podle nich ohrožuje zfalšování a chybná identifikace (*misidentification*). Fyzický objekt je však těžší zfalšovat, než jej vytvořit a kurátoři paměťových institucí se mohou spoléhat na osvědčené postupy pro ověřování autenticity získávaných fyzických dokumentů (např. jejich pořizování ze spolehlivých zdrojů nebo konzultace s odborníky), přičemž jim pomáhají též fyzické vlastnosti originálu. Zfalšování digitálního dokumentu je mnohem jednodušší a předmětem falšování mohou být též jeho metadata a identifikátory. Běžným problémem je rovněž existence různých online verzí dokumentu, které lze těžko odlišit, pokud změny nejsou zdokumentované. Autoři se také zabývají otázkou originálu a jeho surogátu (mikrofilm, digitalizovaný dokument). Zničení originálu je společensky citlivé téma (uvádějí případy pobouření britské veřejnosti, když knihovny odstranily rozpadající se periodika a nahradily je mikrofilmy), ale ekonomicky i prakticky (degradace) nevyhnutelné. Vytvoření surogátu podle autorů nikdy nemůže zachovat *všechno* o objektu, ale nevytvořit žádný surogát by mohlo znamenat (v případě křehkých nebo degradujících originálů), že *všechno* bude ztraceno.

Směrnice NISO (30) uvádí tři sady principů pro budování kvalitních digitálních sbírek (principy pro fondy, objekty fondů a metadata). Jeden z principů zní: „Kvalitní objekt lze autentifikovat (*authenticated*)“. (30, s. 55). Autenticita se týká „míry důvěry (*confidence*), kterou může mít uživatel v integritu (*integrity*) a důvěryhodnost (*trustworthiness*) nějakého objektu“; autentifikace (*authentication*) je definována jako „úkon, kterým se zjišťuje, zda je objekt v souladu se svým zdokumentovaným původem, strukturou a historií a jestli nebyl porušen nebo neautorizovaně změněn.“ (30, s. 55). Také tento zdroj obsahuje zmínku o rozdílech zjišťování autenticity fyzických a digitálních objektů: pro fyzický objekt bylo

možno užít forenzní analýzu; pro digitální objekt je nutno užít jiné metody: dokumentaci digitální provenience, digitální vodoznak a kontrolu neporušenosti (*fixity checking*). Digitální provenience objektu je zde chápána jako původ objektu i historie jeho změn. Informace o původu objektu lze zaznamenat interně (např. do hlavičky souboru), historii změn externě (doporučuje se zápis hodnot elementů událostí PREMIS do schématu METS, konkrétně sekce digiProvMD). Digitální vodoznak lze také užít pro záznam informací o původu (informace o vlastnictví nebo copyrightu). Kontrola neporušenosti je srovnání výstupů algoritmu hašovací funkce vytvořených z objektu v různých časových okamžicích, přičemž předchozí musí být uložen. Podle směrnice může vodítka k autenticitě poskytovat také kontext: autentický objekt by měl mít vyjádřené vztahy s jinými verzemi téhož objektu, jinými objekty sbírky nebo objekty, které tvoří součást tohoto objektu.

Směrnice UNESCO/PERSIST Guidelines for the selection of digital heritage for long-term preservation vydaná v březnu 2016 se zabývá otázkou zachrany digitálního dědictví. Digitální dědictví vyžaduje „aktivní archivační přístupy, aby byla zajištěna jeho autenticita, zpřístupnitelnost (*accessibility*) a použitelnost (*usability*) napříč časem“. (82, s. 16). Autenticitou se však směrnice zabývá okrajově. V terminologickém slovníčku se uvádí, že autenticita (digitálního dědictví) „souvisí s důvěryhodností (*trustworthiness*) archiválie nebo jednotky, tj. s vlastností být tím, za co se vydává (*quality of being what it purports to be*), buď jako originál nebo jako spolehlivá (*reliable*) kopie odvozená z originálu plně zdokumentovanými procesy.“ (82, s. 16). Směrnice se dále zmiňuje o autenticitě metadat: k zajištění jejich autenticity a spolehlivosti (*reliability*) jsou zapotřebí tzv. „metametadata“, což jsou údaje o zdroji a způsobu sestavení metadat (kdy byla metadata sestavena a kým; zda byla získána automatizovaně, nebo manuálně; jaké nástroje a metody byly užity pro jejich sestavení) (82, s. 15).

#### **5.5.1.1 Autenticita a přístupy k uchovávání**

V oblasti digitální archivace uvádí norma dva základní přístupy k uchovávání: digitální migrace a emulace (viz oddíl 2.2.4). V praxi se v rámci prvního přístupu zdůrazňuje zejména formátová konverze. Otázka, jaký přístup má být upřednostňován, byla předmětem odborných diskuzí knihovnické komunity zejména v 90. letech 20. století. Emulace v teoretické rovině znamená zachovat napříč časem původní digitální objekt CDO beze změny. Na objektu CDO se vykonává pouze bitová ochrana (renovace a replikace) a informační obsah je reprodukován emulačními technikami, které umožňují reprodukci informačního obsahu i v budoucím prostředí, kdy již nejsou dostupné aplikace schopné

pracovat s formátem objektu. Mění se tedy pouze interpretační informace, nikoliv objekt CDO. Formátová konverze znamená, že se v případě zastarávání formátů provádí převod objektu CDO do nového formátu, čímž se původní digitální objekt mění.

Nejznámějším zastáncem emulace je počítačový vědec Jeff Rothenberg. V řadě studií publikovaných v 90. letech 20. století argumentoval tím, že pouze emulace zajišťuje největší realizovatelnou záruku autenticity (nezasahování do původního objektu CDO). S kritikou tohoto konceptu přišel teoretik archivní vědy David Bearman, který argumentoval tím, že emulace je nepochopení podstaty uchovávání – k udržení autenticity objektu nestačí uchovat digitální objekt. Je nutné také udržovat související informace o jeho obsahu, struktuře a kontextu tohoto objektu (1, s. 125-126). Otázkami emulace se v disertační práci nebudeme zabývat vzhledem k tomu, že v současné praxi převažuje důraz na formátovou konverzi jako plánové opatření dlouhodobého uchovávání. Obecně je problémem emulace její praktická uskutečnitelnost v dlouhodobém horizontu. V současné praxi je také uznáváno, že formátová konverze může ohrozit autenticitu, zejména v případě nevratné konverze.

### **5.5.2 Autenticita a zpřístupnění**

Směrnice MINERVA tematizují autenticitu v souvislosti s prezentováním výstupu digitalizačního projektu veřejnosti: „Název domény tvoří část „obchodní značky“ projektu a napomůže koncovým uživatelům určit autenticitu poskytovaného obsahu. Názvy domén by tedy měly být jasné označeny buď názvem projektu, nebo organizací, která projekt zajišťuje. Pro zvýšení důvěry koncových uživatelů, že si vyměňují informace se správným webovým sídlem projektu, může být někdy vhodné zabezpečit síťové spojení mezi klientem a serverem protokolem Secure Sockets Layer (SSL).“ (28, s. 66). Tato směrnice tedy autenticitu dokumentu pojímá pouze z hlediska jeho původce: důvěryhodnost webového sídla digitální knihovny (tj. skutečnost, že tato digitální knihovna je skutečně provozována původcem, za jehož webové sídlo se vydává) plošně činí zde prezentované digitální dokumenty autentickými.

Standard PREMIS, jak bylo výše uvedeno, otázku autenticity zpřístupňovaných dokumentů nezužuje na otázku původce. V souvislosti s původcem zpřístupňovaného obsahu uvádí užití digitálního podpisu. Archiv může připojit digitální podpis k datům, která vydává jako balíčky DIP, jako doklad tohoto, že tento archiv je skutečně původcem daných dat (52, s. 259). Zde tedy nejde o plošnou záruku autenticity důvěryhodným webovým sídlem, ale otázku autenticity původce spojenou přímo s konkrétními daty. V praxi se spíše bude jednat o situaci, kdy archiv vydává balíčky DIP jiné organizaci.

Dalším aspektem zmiňovaným v souvislosti s autenticitou v oblasti zpřístupnění je citovanost. Publikace organizace Digital Preservation Coalition uvádí, že autenticita se „v případě nativně digitálních nebo digitalizovaných dokumentů týká skutečnosti, že cokoliv je citováno, je stejné, jako bylo v době svého vzniku, pokud nejsou v doprovodných metadatech uváděny žádné změny.“ (83). Publikace dále uvádí, že metadata a dokumentace zaznamenávající změny dokumentu jsou možná jediným prostředkem pro zjištění autenticity digitálního dokumentu (84).

### **5.5.3 Autenticita a digitalizace**

#### **5.5.3.1 Autenticita a metadata v digitalizaci**

Jak bylo uvedeno v oddíle 5.3, podle normy ISO 14721 jsou provenienční informace součástí dokladů o autenticitě. V tabulce konkrétních příkladů pro oblast digitalizace norma uvádí metadata o digitalizačním procesu a odkaz na obrazovou matici (10, s. 75). Vnímaná důležitost zaznamenání co největšího množství zejména technických metadat během digitalizace již byla zmíněna v přehledu současné praxe, stejně jako skutečnost, že hlavním vodítkem pro tuto činnost jsou metadatové standardy (zejména PREMIS a MIX), které nabízejí vymezené typy informací, jež lze zaznamenávat. Otázkou je, jaká metadata shromažďovaná v digitalizaci jsou klíčová z hlediska autenticity a jaká již nikoliv. Grycz jako příklady technických metadat procesu fotografování, která považuje za důležitá z hlediska autenticity digitalizátu, uvádí: typ fotoaparátu, datum snímání, nastavení apertury a expozice, bitová hloubka a hodnoty rozlišení (62, s. 47).

#### **5.5.3.2 Autenticita a koncept věrnosti**

Zdaleka nejrozšířenějším konceptem objevujícím se v literatuře věnované digitalizaci je věrnost (*faithfulness*). Věrnost bývá v některých případech spojována právě s konceptem autenticity. Za zřejmě nejvýznamnější zdroj pro koncept věrnosti digitalizátů lze považovat směrnici americké Federace digitálních knihoven (Digital Library Federation; DLF). Směrnice MINERVA (28) pojem věrnost vůbec neuvádějí, směrnice FADGI se v otázce věrnosti právě na směrnici DLF odkazují (20, s. 51).

Směrnice DLF definují věrné digitální kopie (*faithful digital reproduction*) jako „digitální objekty, které jsou optimálně formátovány a popsány s ohledem na jejich kvalitu (funkčnost a užitná hodnota), trvalost (dlouhodobé zpřístupňování) a interoperabilitu (např. napříč platformami a softwarovými prostředími)“. Cílem věrnosti je „přesně reprodukovat výchozí zdrojový dokument, s ohledem na jeho úplnost, vzhled původních

stránek (včetně tonality a barvy) a správnou (tj. původní) posloupnost stránek.“ (85, s. 2). Ke splnění tohoto cíle uvádějí několik podmínek (85, s. 4):

- Obrazy vakátů (včetně zadní strany desek) budou začleněny jako části stanovené posloupnosti.
- Bude možné propojit obecná deskriptivní metadata s digitálními součástmi objektu (např. pro potřeby citace).
- Bude možné z digitálních kopií reprodukovat (tiskem nebo online zobrazením) původní posloupnost částí, ale rovněž navigaci mezi nimi (zobrazit následující stranu, první apod.), bude možné je citovat.

Pouze kopie, které splňují tato kritéria, lze podle směrnice označit za věrné. Pojem autenticita se zde nevyskytuje. V některých jiných zdrojích je však koncept autenticity spojen právě s věrností digitalizátů (viz např. [ (86); (62)]); někdy je tato souvislost implicitní. Například výraz „autenticita“ se ve směrnici FADGI vyskytuje jednou, v pasáži zabývající se problémem skenování reliéfních pečetí, které většina zařízení nenasnímá, což podle směrnice „vyvolává otázky o autenticitě digitální reprezentace dokumentů.“ (20, s. 51). Z kontextu je zřejmé, že se mluví o věrnosti. Explicitní spojení autenticity s věrností digitalizace nalezneme v popisu projektu (zaměřeného na digitalizaci rukopisů) Octavo: „Digitální surogát musí zachycovat dostatečné informace, které napomohou badateli (řekněme za 300 let) vidět nebo reprodukovat snímané obrazy způsobem, který představuje autentickou reprezentaci (*authentic representation*) toho, jak originál vypadal v době svého nasnímání.“ (62, s. 47).

Specifickými otázkami věrnosti dále jsou: jaký strukturální model digitální reprezentace knihy je věrným, jak zajistit barevnou věrnost a jaké úpravy původních snímků provádět.

Grycz ve svém příspěvku ve sborníku věnovaném vytváření, uchovávání a zpřístupňování digitálního dědictví představuje následující přístup k těmto otázkám. Tento přístup se sice vztahuje zejména k digitalizaci, nicméně je pro účely této disertační práce zajímavý přinejmenším jako ukázka možnosti odlišného přístupu k digitalizaci. Grycz uvádí: „Abychom zůstali věrni tradici výroby knih, měly by digitální obrazy vzácných tisků reprodukovat časem ověřené kvality původního (original) tištěného díla... Jedním z nejpatrnějších způsobů [jak tohoto docílit] je trvat na digitalizování knihy jako otevřené dvojstránky, bez ořezaných hran, tedy tak, aby čtenář mohl knihu vnímat, jako by spočívala



na stole nebo byla držena v rukou.“ (62, s. 41). Takto nasnímané předlohy podle něj také umožňují zachycení informací, které přesahují na obě stránky. Grycz tento model reprezentace označuje jako snímání „v přirozeném stavu, který je věrný způsobu, jakým byla původní kniha navržena k prohlížení a užívání“ (62, s. 42).

Pro zajištění obrazové věrnosti prezentuje Grycz přístup, kdy je snímána barevná kalibrační tabulka a palcové nebo metrické pravítko spolu s každým svazkem; pro každý projekt je pak vytvářen vlastní ICC profil (62, s. 47). Užití ICC profilu je podle něj způsobem, jak snížit riziko ztráty barevné věrnosti v budoucnosti: „Dostupnost (barevných) profilů může přispět k vyšší důvěře, že snímek bude zobrazitelný a bude barevně věrný po celé generace. Lze očekávat, že v budoucnosti budou dostupná zařízení a procesy zcela odlišné od současně užívaných. Pokud však nová zařízení budou rovněž odpovídat převedeným ICC standardům, pak bude pravděpodobnost přesného zobrazení a reprodukce vyšší.“ (62, s. 48).

Směrnice FADGI za účelem dosažení stejného cíle (zachování možnosti barevné věrnosti i v budoucím technologickém prostředí) doporučují převést barevné profily závislé na zařízeních do barevného profilu Adobe RGB 1998 (buď nastavením na skeneru, nebo přiřazením tohoto profilu do obrazových dat během zpracování) (20, s. 48). Tyto směrnice, stejně jako Grycz, nedoporučují provádět žádné ořezy s tím, že „kolem celého dokumentu by měly být viditelné malé okraje.“ (20, s. 50).

#### **5.5.4 Autenticita původních snímků**

V oblasti profesionální fotografie je otázka dodatečných obrazových úprav původních snímků předmětem odborných polemik, které se dále liší podle kontextu užití (umělecká, reklamní nebo dokumentární fotografie). Nalezneme například následující tematizaci: „Digitální díla narušují naše pojetí autentické reprezentace, zejména ve fotografii, kde nejde již o pouhé retušování, ale vytváření zavádějících reprezentací. To narušuje naši víru ve vizuální dokumentaci – v některých případech se digitální manipulace přímo očekává.“ (86). Uvedená tematizace pak bývá přebírána i do kontextu digitalizace tištěných dokumentů, kdy rozšířeným přístupem je původní snímky obrazově neupravovat vůbec, na základě přesvědčení, že právě tyto snímky představují nejvěrnější digitální reprezentaci tištěného dokumentu.<sup>63</sup> Podle směrnice FADGI je však v digitalizaci toto přesvědčení zavádějící, a to vzhledem k tomu, že sama snímací zařízení již v procesu snímání (tj. ještě před uložením

---

<sup>63</sup> Tento přístup výstižně shrnuje následující tvrzení: „Naskenovaná digitální reprezentace by měla být co nejméně digitálně modifikována tak, aby výsledná digitální reprezentace věrně reprezentovala původní analogový dokument.“ (122, s. 139)

obrazových dat snímkového formátu) provádějí širokou škálu obrazových úprav. Tento přístup podle směrnic také předpokládá, že se dokážeme vypořádat s nedostatky snímacího zařízení lépe než výrobce a že nejsme omezeni časem (20, s. 39). Zachování původních snímků bez úprav směrnice doporučují pouze tehdy, pokud snímky splňují požadované parametry. V ostatních případech je podle nich vhodné provádět menší úpravy pro optimalizaci obrazové kvality, což je mj. přínosné z hlediska možnosti zacházet se všemi soubory stejným způsobem.

## 5.6 Další vybrané tematizace autenticity

### 5.6.1 Benjaminova koncepce jedinečného originálu

S tematizací autenticity se často setkáváme v oblasti tradičního výtvarného umění. Autenticita zde bývá neodlučně spojována s konceptem originálu a jedinečnosti.<sup>64</sup> Jedním ze základních zdrojů těchto přístupů je studie Umělecké dílo ve věku své technické reprodukovatelnosti (Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit) Waltera Benjamina z roku 1936 (87, s. 17-47). Tento známý německý filozof a estetik v ní podrobuje kritice seriálovou produkci moderní doby. Hlavní princip této produkce, technická reprodukovatelnost objektu, podle něj představuje nový společenský jev, který nemá v dějinách obdoby: „Umělecké dílo bylo v zásadě vždy reprodukovatelné. Co lidé udělali, to mohli také pokaždé napodobit. Takové kopírování prováděli žáci, aby se v umění cvičili, mistři, aby šířili svá díla, a do třetice ti, kteří bažili po zisku.... Řekové znali pouze dva způsoby technické reprodukce uměleckých děl: lití a ražení. Bronzy, terakoty a mince byly jedinými uměleckými díly, která dokázali vyrábět masově. Všechna ostatní díla byla jedinečná a technicky nereprodukovatelná.“ (87, s. 17-18).

Problém spatřuje Benjamin v tom, že i sebedokonalejší technické reprodukce postrádají jeden prvek – přítomnost výtvarného díla v čase a prostoru. Jedinečná (technicky nereprodukovatelná) díla podle něj disponují původní časoprostorovou konkrétností, kterou nazývá „zde a nyní“ díla: Toto „zde a nyní“ je vlastní pouze originálu a „vytváří pojem jeho pravosti (*Echtheit*). Chemické analýzy patiny na bronzu mohou napomoci ke stanovení pravosti; podobně jako k tomuto může vést důkaz, že určitý středověký rukopis pochází z archivu patnáctého století.“ (87, s. 3).

---

<sup>64</sup> Srv. výzkum užití pojmu autenticita v oblasti umění (116, s. 15-16).

Při technické reprodukci originálu mizí nejen jeho originální časoprostorové ukotvení, ale i původní kontext užití. Kopii je možno užívat způsobem, který nebyl možný v případě originálu: „Katedrála opouští své místo, aby našla umístění v pracovně milovníka umění; sborové dílo, provedené v sále či pod otevřeným nebem, lze poslouchat v pokoji.“ (87, s. 18). To, co kopie ztrácí, nazývá „aurou“: „Reprodukční technika... vyděluje reprodukované dílo z dosahu tradice. Zmnožováním reprodukce klade na místo jedinečného výskytu tohoto díla výskyt masový. A tím, že reprodukci dovoluje, aby vyšla vstříc vnímateli v jeho nynější situaci, aktualizuje reprodukované“ (87, s. 4). Možnost reprodukovatelnosti uměleckých děl (i jiných objektů) masovými médii je tak podle Benjamina novou etapou v dějinách lidstva, kterou charakterizuje nekonečný proces kopírování. Technologické prostředky dovolují donekonečna rozmnožovat cokoliv, čímž zaniká originalita a nastává „seriálové mlčení“ (vytržení z kontextu, ztráta autority věci).

Pojem autenticity zmiňuje Benjamin pouze okrajově: uvádí, že kdysi měla výtvarná díla kultovní hodnotu, v moderní době „se sekularizací umění nastupuje autenticita (*Authentizität*) na místo kultovních hodnot“ (87, s. 6). Tato studie je však důležitá jako ukázka rozšířeného pojetí, podle kterého je klíčovou vlastností nejocetňovanějších (nedigitálních) výtvarných děl to, že jsou originálem, který je jedinečný. Jedinečný originál je neodlučně spjat s konkrétní materií, a tak existuje pouze v jednom „exempláři“ a žádná jeho kopie nemůže dosáhnout hodnoty originálu. Studie je však zajímavá i v jiném ohledu, a sice jako předzvěst dokonalého digitálního kopírování, kdy otázka materie přestává být z principu podstatná i pro kopie.

### 5.6.2 Autenticita ve směrnici DSA

Směrnice DSA (Data Seal of Approval) byla původně vytvořena pro posuzování důvěryhodnosti vědeckých archivů spravujících výzkumná data. Jejím sedmým požadavkem je zaručení integrity a autenticity dat. Jako vysvětlující text se uvádí: „Integrita zaručuje, že změny dat a metadat jsou zdokumentovány a lze dosledovat jejich zdůvodnění a původce. Autenticita se týká míry spolehlivosti původně vložených dat a jejich provenience, včetně vztahu mezi původními daty a zpřístupňovanými daty, a toho, zda jsou, nebo nejsou udržovány vztahy mezi datovými sadami a metadaty“ (88, s. 18) Směrnice dále uvádějí následující sady dokladů (*evidence*), které musí dodat archiv certifikujícímu subjektu.

Doklady integrity jsou: popis kontrol, které ověřují, zda digitální objekt nebyl pozměněn nebo porušen (např. kontroly neporušenosti); dokumentace úplnosti dat i metadat;

podrobnosti o tom, jak jsou všechny změny dat a metadat zaznamenávány; popis strategie řízení verzí; užití odpovídajících mezinárodních norem a zvyklostí a jejich popis.

Doklady o správě autenticity (*authenticity management*) se vztahují k těmto otázkám:

- Má archiv strategii pro změny dat? Jsou producenti srozuměni s touto strategií?
- Udržuje archiv data o provenienci a související auditní záznamy?
- Udržuje archiv odkazy na metadata nebo jiné datové sady? Pokud ano, jakým způsobem?
- Srovnává archiv nezbytné (*essential*) vlastnosti různých verzí téhož souboru? Jakým způsobem?
- Kontroluje archiv identitu vkladatelů?

Směrnice DSA dále uvádějí, že tento požadavek se týká celého životního cyklu dat v archivu a vztahuje se k pracovním postupům souvisejícím s příjmem dat a jejich posuzováním, k dokumentaci úložných procedur, k plánu pro uchovávání, k otázkám vyhledávání a k identifikaci a opětovnému využití dat.

### 5.6.3 Autenticita zdrojů výzkumného procesu

David Bearman a Jeniffer Trantová publikovali v roce 1998 studii, která se zabývá dopady digitalizace dokumentů a zpřístupňování digitálně nativních i digitalizovaných dokumentů v internetové síti na autenticitu zdrojů užitých při výzkumném procesu. Posuzování autenticity musí být podle autorů založeno na „posouzení původu, úplnosti a vnitřní integrity dokumentu“ a může také „vycházet z konzistence a koherence, která existuje mezi konkrétním zdrojem a jinými zdroji stejného typu nebo ve stejném kontextu.“ (86).

Obavy z dopadů technologických změn na autenticitu dokumentů nejsou podle autorů novým jevem (obavy z vynálezu knihtisku daly kdysi vzniknout klasické diplomacie). Otázky padělání fyzických dokumentů podle nich nikdy nebyly v centru zájmu vědců, a to z důvodů technické náročnosti vytvoření přesvědčivých padělků a obtížného vstupu takových padělků do oblasti zdrojů považovaných za směrodatné mezi vědci. S nástupem digitalizace a zpřístupňování dokumentů na internetu se však objevují nová rizika, která výzkumný proces znesnadňují. Bearman a Trantová mezi ně řadí zejména pravděpodobnost existence různých verzí digitálně nativních dokumentů, které lze od sebe obtížně odlišit,

a nedostatečnou znalost o tom, kým a jakým způsobem byly digitalizační technologie užity při vytváření zdrojů.

Je třeba dodat, že hlavním zájmem autorů z hlediska digitalizovaných zdrojů jsou především jedinečné fyzické originály (tedy např. rukopisy), ke kterým se výzkumník může dostat jen díky internetu. Tato nová možnost na jednu stranu umožňuje výzkum tam, kde předtím nebyl možný (např. pro historiky zkoumající rukopisy), na druhou stranu přináší problém existence různých digitálních reprezentací téhož rukopisu nebo jiného typu fyzického originálu, které se od sebe liší (např. barevností), což podle autorů vede k tomu, že „všechny otázky týkající se autenticity originálu jsou převáženy dodatečnými otázkami způsobů reprezentace.“ (86). Požadavky na autenticitu podle nich souvisejí s požadavky samotného procesu výzkumu. Vymezují tři fáze procesu: vyhledávání, získání a užití informačních zdrojů.

Ve fázi vyhledávání je klíčová jednoznačná identifikace digitálně nativního dokumentu, která dokáže odlišit různé verze nebo vydání. Autoři jako řešení uvádějí užití perzistentních identifikátorů důvěryhodných systémů, například identifikátoru DOI. Druhý problém pak představují různé digitální reprezentace téhož fyzického zdroje (tedy digitalizované dokumenty vzniklé v různých projektech a institucích různými způsoby digitalizace). Pro posuzování autenticity digitalizovaného zdroje je podle autorů důležité, aby digitální reprezentace byla transparentní, tj. aby bylo možné zjistit, jakým konkrétním způsobem byla vytvořena. Roli při tomto posuzování také hraje znalost původce (organizace, která digitalizát vytvořila) a jeho důvěryhodnost.

Ve fázi získání (tj. stažení) dokumentu jde o otázku toho, nakolik může výzkumník důvěřovat tomu, že digitální dokument je skutečně tím, za co se vydává. Zde podle autorů hrají roli metody, jako jsou hašovací funkce, digitální vodoznak nebo šifrování. Problémem však mohou být změny, které nastanou po stažení digitálního dokumentu do počítače výzkumníka (např. v digitální knihovně mohou být dostupné funkce, které po stažení nebudou na počítači výzkumníka dostupné). Ve fázi užití jsou podle autorů klíčová přidružená metadata a dokumentace, zahrnující informace o technologii vytvoření dokumentu.

Autoři v závěru volají po potřebě stanovit požadavky autenticity pro konkrétní oblasti: „Měli bychom lépe porozumět odlišným požadavkům na autenticitu, a ne hledat jedno jediné řešení.“ (86).

#### 5.6.4 Rothenberg a autenticita spojená s vhodností

Rothenberg se ve své konceptualizaci autenticity z roku 2000 zabývá autenticitou „informační entity“ (*informational entity*), což je „entita, jejíž účel nebo úloha je informační“ (89, s. 52). Součástí digitální informační entity je i software a hardware nutný k její reprodukci. Rothenberg v úvodu uvádí: „Autenticita a technologické otázky obklopující uchovávání digitálních informačních entit se vzájemně ovlivňují způsobem, který je nový a dalekosáhlý. Je mnohem pravděpodobnější, že důsledkům těchto interakcí lépe porozumíme, pokud vytvoříme jednotný, koherentní a oborově nezávislý pohled na autenticitu.“ (89, s. 51). Rothenberg hledá univerzální model autenticity, na základě něhož by jednotlivé obory mohly určovat specifické sady požadavků.

Sám pojem autenticity vymezuje poměrně široce: autenticita zahrnuje podle něj „otázky integrity, úplnosti, bezchybnosti (*correctness*), validity (*validity*), věrnosti (*faithfulness*) originálu, smysluplnosti (*meaningfulness*) a vhodnosti (*suitability*) pro zamýšlený účel“ (89, s. 52). V uchovávání se pak „ve své podstatě vztahuje k času.“ (89, s. 58). Konkrétnější definice se podle něj budou lišit podle různých oborů.

Rothenberg kritizuje běžný způsob určování autenticity, který nazývá „strategie původnosti“ (*originality strategy*). Tato strategie podle něj funguje dobře pro tradiční entity, ale je problematická pro digitální entity. Strategií původnosti je „zaměřit se na původnost informační entity, tj. zda zůstala nezměněna vzhledem ke svému původnímu (*original*) stavu“ (89, s.56). K tomu se podle něj využívají dvě metody (*tactics*).

První metodou je „zaměřit se na vnitřní vlastnosti (*intrinsic properties*) nějaké informační entity a poskytnout kritéria, podle kterých se určí, zda je každá vlastnost přítomná ve své náležitosti, původní podobě“ (89, s. 56). Pokud například stanovíme, že těmito vlastnostmi jsou papír a inkoust, pak můžeme užít chemických testů k ověření toho, zda jsou papír a inkoust původní, či nikoliv, a na základě toho stanovit, zda je entita autentická.

Druhou taktikou je zaměřit se na proces, kterým je informační entita uchovávána. V tomto případě je doklad historie správy zárukou toho, že entita nebyla změněna, nahrazena nebo poškozena, a je tedy autentická. Tuto taktiku podle něj užívá archivnictví: poskytuje dokumentaci o původu, původci a kontextu vzniku archiválií a dokládá, že archiválie byly udržovány nepřetržitým řetězcem správy, během něhož nebyly porušeny. Archivnictví je tak podle Rothenberga založeno na dvou podmínkách: 1) musí být udržen nepřerušovaný řetězec správy archiválie; 2) během správy nesmí být archiválie nepatřičným způsobem změněna.

Skutečnost, že správa byla vykonávána nepřetržitě, sama o sobě nemůže dokázat, že archiválie nebyla změněna. Jelikož je však obtížné dokázat neporušenost archiválie, slouží doklad nepřerušené správy podle Rothenberga jako náhražka důkazu její neporušenosti (89, s. 56-57).

Rothenberg navrhuje vlastní strategii určování autenticity, kterou nazývá strategii vhodnosti (*suitability strategy*), což znamená „definovat autenticitu z hlediska toho, zda je informační entita vhodná pro určitý účel.“ (89, s. 57). Podle Rothenberga je jím navrhovaná strategie lepší než strategie původnosti, protože předpokládá zachování určitých vlastností, které jsou stanoveny z hlediska účelu, za jakým je entita uchovávána: „autenticita uchovávaných informačních entit je v jakékoliv oblasti v konečném důsledku svázána s jejich vhodností pro konkrétní účely, které jsou předmětem zájmu v dané oblasti“ (89, s. 58). Proto je podle něj potřeba „pokusit se vytvořit principy autenticity (*authenticity principles*) pro různé oblasti nebo obory, které umožní definovat autenticitu z funkčního hlediska. Princip autenticity shrnuje celkový záměr autentického uchovávání z konkrétních právních, etických, historických, uměleckých nebo jiných hledisek.“ (89, s. 59). Pro každý princip je pak užitečné odvodit sadu kritérií autenticity (*authenticity criteria*) pro konkrétní způsob užití entity, na základě kterých se vytvořit konkrétní požadavky na uchovávání a konceptuální a praktické testy úspěšnosti konkrétních archivačních technik. Za tímto účelem je vhodné popsat škálu očekávaných možností užití. Rothenberg uznává, že takový popis je v principu spekulativní (předpokládá budoucí užití), ale je nutné se o něj pokusit.

Rothenberg se dále zabývá otázkou originálu. Digitální informační entity jsou logicky nezávislé na datových nosičích a digitální entita nemůže existovat jako jedinečná věc (jako je fyzický originál např. obrazu). Pojem originál však chce Rothenberg, vzhledem k jeho všudypřítomnosti v konceptualizacích, zachovat. Digitální originál je podle něj „jakákoliv reprezentace digitální informační entity, u níž existuje nejvyšší možná pravděpodobnost, že zachovává všechny smysluplné a relevantní aspekty této entity“ (89, s. 66). Neurčitost definice (spojení „nejvyšší možná pravděpodobnost“) podle něj nutně vychází z toho, že je velmi obtížné přesně specifikovat všechny tyto aspekty, a také z toho, že je fyzicky a logicky nemožné, aby digitální originál udržoval všechny tyto aspekty, podobně jako to nelze zaručit u fyzických entit (z důvodů přirozené degradace).

Jako příklad možných stupňů požadavků na reprezentace vytvořené na základě takto pojatého originálu uvádí následující úrovně (od nejvyšší míry zachování po nejnižší) (89, s. 62):

- zachování všech předpokládaných záměrů a účelů;
- zachování stejných funkcí a vztahů k jiným informačním entitám;
- zachování stejného vzhledu („look and feel“);
- zachování stejného obsahu (jakkoliv definovaného);
- zachování stejného popisu (tj. zachování pouze metadat o entitě).

Podle Rothenberga jeho definice originálu nepředpokládá žádný konkrétní technický přístup k uchovávání, ale emulace podle něj může být vhodným řešením.

#### **5.6.5 Gladney a struktura tvrzení o autenticitě**

Gladney v kapitole své knihy věnované autenticitě uvádí, že skutečnost, že adjektivum „autentický“ se užívá pro popis velmi odlišných tříd entit, naznačuje sdílený pojmový základ: „jakékoliv tvrzení o autenticitě srovnává nějakou entitu zde a nyní s něčím, co existovalo předtím.“ (90, s. 97).

Tvrzení, že něco je autentické, je podle něj složeno z následující částí (90, s. 97):

- Tvrzení o odvození (*derivation statement*):

„Objekt V je kopií objektu Y.“

- Tvrzení o provenienci (*provenance statement*):

„Tvůrce X vytvořil objekt Y jako součást události Z.“

Podle Gladneyho platí, že „objekt V je autentickou kopií objektu Y, pokud je dostatečně věrnou odvozeninou s pravou proveniencí.“ (90, s. 97). Pravá provenience podle něj znamená, že tvrzení o odvození a tvrzení o provenienci jsou pravdivá. Dostatečná věrnost znamená, že toto odvození odpovídá sociálním konvencím pro daný typ objektu a daný kontext. A shrnuje: „Nejsme ochotni přijmout, že objekt V je autentický, pokud není explicitně určena minulé existence objektu Y, z něhož byl vytvořen objekt V, identita tvůrce objektu Y a konkrétní historická událost Z, v rámci níž k tomuto vytvoření došlo.“ (90, s. 97).



Gladney upozorňuje, že každé odvození přidává, ubírá nebo mění informace nesené původním objektem. Jeho model se vztahuje na všechny objekty, včetně materiálních artefaktů, neboť užívá pojem „kopie“ ve specifickém smyslu: „V každodenním užití myslíme slovem „autentický“ něco jiného, pokud zmiňujeme slavný obraz nebo židli Ludvíka XV., než popisujeme signály.... Nemůžeme udělat nic lepšího než srovnat současný artefakt s jiným objektem. Tímto jiným objektem by mohl být buď konceptuální objekt (to, jak si představujeme, že artefakt vypadal někdy v minulosti), nebo informační objekt, který reprezentuje to, čím byl artefakt v té době. V tomto smyslu můžeme říci, že židle Ludvíka XV. existující v současnosti je „kopií“ židle Ludvíka XV. z doby před 300 lety.“ (90, s. 101).

Gladney dále uvádí, že požadavky na autenticitu se budou lišit podle informačních žánrů (typů dokumentů a jejich kontextů) a typů reprezentace (např. digitální kopie, tištěný list): „V případě digitálních a analogových signálů - informačních přenosů – je otázka autenticity vždy otázkou autenticity repliky. Pro materiální objekty je obvykle otázkou srovnání mezi nějakým předešlým stavem tohoto objektu. Pro knihy a nahrávky muzických umění se rozlišuje mezi informací zapsanou na nějakém materiálním substrátu a samotným substrátem samotným“ (90, s.97).

## **5.7 Shrnutí a dílčí závěr**

Ve zkoumaných odborných zdrojích se tematizace autenticity zabývají alespoň jedním z uvedených okruhů: definice autenticity (podstata autenticity, její předmět a složky apod.), udržení autenticity (předpoklady, podmínky, omezení, postupy atd. pro udržení / zachování autenticity objektu v jeho životním cyklu) nebo ověření / posouzení autenticity (jakým způsobem / do jaké míry lze ověřit / posoudit, že datový objekt je autentický). Hlavním zdrojem pro popsání tematizace autenticity v oblasti digitálního dědictví je archivní teorie, odkud je přebírána klasická definice, podle níž je autenticita objektu vlastností tohoto objektu, že tento objekt je tím, za co se vydává, a který nebyl zfalšován nebo změněn. Dva uvedené archivářské projekty přinesly první konceptualizaci autenticity určitého typu digitálního dokumentu (elektronické archiválie). Oba jsou založeny na postupu, kdy je nejprve vytvořen model pro elektronickou archiválii jako model předmětu autenticity, a následně jsou zkoumány jeho implikace pro otázky autenticity.

Projekt InterPARES přišel s paradigmatickým posunem v kontextu uchovávání v archivnictví, podle kterého nelze uchovávat elektronické archiválie v podobě, v jaké jsou vnímány příjemcem; je možné uchovávat pouze reprodukovatelnost uložených bitů

do vnímatelné podoby. Typy bitových toků tvořících archiválii, které vyžadují odlišné technologické zacházení, projekt označil jako digitální komponenty. Autenticitu archiválie rozdělil na dvě složky: identitu (atributy, které ji jednoznačně charakterizují a odlišují od jiných archiválií) a integritu (úplnost a neporušenost digitálních komponent). Projekt připustil možnost bitových změn digitálních komponent nenarušujících integritu archiválie. Musí být však splněno několik podmínek, především musí být zachováno sdělení, které archiválie přenáší, a udrženy požadované znaky dokumentové formy. Archiválie jsou ohroženy kdykoliv při procesu vytváření kopií. Posoudit autenticitu archiválie znamená zjistit její identitu a prokázat její integritu, což znamená podložit tvrzení o autenticitě dodatečnými doklady. Za tímto účelem projekt InterPARES stanovil dvě sady požadavků autenticity: pro posuzování autenticity archiválií přijímaných do archivu a pro následné udržování jejich autenticity během uchovávání v archivu. Tyto požadavky jsou částečně specifické pro kontext archivnictví, částečně je lze aplikovat i na jiné typy dokumentů. Klíčovým prvkem dokladů autenticity je dokumentace vytvářená archivem, zejména popis způsobů vytváření kopií. Celkově lze říci, že konceptualizace projektu zůstává v obecnější rovině: nezabývá se specifickými požadavky odlišných typů dat (textová, obrazová apod.), neproblematizuje otázku reprodukce digitálních komponent archiválie, otázku formátů a je určena pro správu archiválií v běžných systémech pro správu dokumentů, jejichž soulad s funkčními požadavky normy ISO 14721 je sporný.

Řada požadavků pro udržování autenticity archiválií projektu InterPARES byla již mnohem podrobněji rozpracována ve funkčním modelu OAIS prvního vydání této normy. Norma ISO 14721:2012 však napravila nedostatek předchozího vydání a nově definuje pojem autenticita. Uvádí pouze obecný rámec pro požadavky na autenticitu: definici pojmu a jeho vztah ke dvěma dalším pojmům. Klade důraz na to, že autenticita je otázkou míry a je posuzována na základě dokladů, a nově přidává do definice pojmu dlouhodobé uchovávání požadavek uchovávat spolu s informacemi také doklady o jejich autenticitě. Součástí těchto dokladů jsou provenienční informace, které specificky v případě nevratných transformací musejí zahrnovat také popis transformačních vlastností informací (toho, co má být zachováno).

Standard PREMIS ve své vysvětlovací části uvádí vlastní tematizaci autenticity, která převádí obecnější koncepty do podoby konkrétních elementů. Za klíčovou součástí autenticity považuje záznam provenienčních informací, zejména o odvozování (kopírování, transformace) a o signifikantních vlastnostech (jako vlastnostech, které mají být zachovány napříč odvozováním). Do souvislosti s autenticitou dává i otázky neporušenosti souborů

a mechanismů jejího ověření a integrity souboru, kterou vztahuje k formátové identifikaci a validaci. Tato tematizace se pak odráží v konkrétních metadatových elementech a předepsaném způsobu jejich užití.

Představené tematizace autenticity knihovnictví se zaměřují jen na některé dílčí aspekty nebo části životního cyklu digitálních dokumentů z hlediska autenticity. Definice zde užívané jsou vesměs převzaty z archivnictví. Nejrozšířenější tematizací, která je někdy dávána právě do vztahu s autenticitou, je tematizace věrnosti digitalizačního převodu (věrnost jako autentická digitální reprezentace tištěné předlohy). Zajímavá je i tematizace autenticity zpřístupňovaných dokumentů obsahující otázky důvěryhodnosti webového sídla digitální knihovny nebo citační praxe. Jediným společným prvkem všech tematizací z této oblasti je zdůraznění vztahu mezi autenticitou a metadaty nebo dokumentací (specificky např. požadavek zaznamenávat všechny změny dokumentu).

Článek Bearmana a Trantové z oblasti výzkumných procesů obsahuje důležitý poukaz na potřeby perzistentní identifikace nezávislé na URL adrese, která umožňuje odlišení různých verzí dokumentu, a transparentní dokumentace digitalizačního procesu umožňují posoudit kvalitu digitalizátu. Uvedené texty Rothenberga a Gladneyho představují specifické konceptualizace autenticity, jejichž definiční preciznost (v oblasti knihovnictví málo vídaná) odkazuje k erudici počítačových vědců. Rothenbergův koncept strategie vhodnosti je významný především tím, že poukazuje na potřebu zohlednit požadavky na autenticitu v různých oblastech, a to z hlediska vhodnosti dokumentu pro určitý účel. Autor upozorňuje, že v různých kontextech mohou být důležité jiné vlastnosti nebo aspekty digitálního dokumentu, a doporučuje pojímat objekt autenticity jako soubor vlastností požadovaných pro určitý účel. Gladney také upozorňuje na odlišné požadavky pro jiné typy objektů. Jako základní aspekty autenticity identifikuje aspekt provenience a odvozeniny a otázku věrnosti považuje za otázku kontextu.

Na základě přehledu tematizací autenticity byly zjištěny následující klíčové kategorie pro zkoumání autenticity: model pro předmět autenticity; časoprostorový aspekt; otázka originálu; aspekt odvozeniny; koncept věrnosti digitalizace; klíčové vlastnosti; identita a integrita dokumentu; koncept provenience; dokumentace a metadata; problémy reprodukce; rizika pro udržení autenticity. Tyto kategorie tak budou rámcem pro revizi uvedených tematizací v následující kapitole.

## 6 Revize konceptů autenticity

Cílem této kapitoly je provést revizi tematizací autenticity představených v předchozí kapitole a na základě ní vytvořit obecný rámec pro zkoumání specifických požadavků autenticity v oblasti řízení životního cyklu digitalizátů knih, který je předmětem kapitoly následující. První část této kapitoly (podkapitola 6.1) popisuje základní kontext analýzy (institucionalizovaná správa a typ dokumentu). Druhá část (podkapitola 6.2) provádí revizi tematizací autenticity uvedených v předcházející kapitole, a to v rámci kategorií, které byly určeny v dílčím závěru předcházející kapitoly. Třetí část (podkapitola 6.3) představuje návrh obecného rámce pro zkoumání specifických požadavků autenticity v kontextu řízení životního cyklu digitalizátů knih, který je založen na výstupech této revize.

### 6.1 Kontext pro zkoumání autenticity

#### 6.1.1 Aspekt institucionalizace

Ve většině uvedených definic je hlavním prvkem autenticity datového objektu jeho vlastnost být tím, za co se vydává, případně s dodatkem, že nebyl zfalšován nebo nepatřičným způsobem změněn (tj. bez dokumentace změny nebo zachování požadovaných vlastností). Otázka, nakolik se objekt sám může za něco vydávat, je nepochybně vždy otázkou kontextu, ve kterém se objekt vyskytuje, sociokulturního prostředí a jím formovaných očekávání. Je zřejmé, že v neobecnější rovině jsou pro účely této disertační práce relevantní pouze objekty, které jsou lidskými výtvoři. V oblasti přírodních objektů může být otázka autenticity čistě otázkou toho, že daný objekt je skutečně přírodním objektem (a tedy ne lidským podvrhem, např. že nalezená kost je skutečně kostí dinosaura, a nikoliv lidským padělkem). V oblasti lidských výtvorů je pak zpřesněním tohoto kontextu skutečnost institucionalizované správy, jedná se tedy o autenticitu dokumentu, který je evidován a uchováván danou organizací. Například na prodej knih v knihkupectví nebo výkup knih antikvariátem se budou vztahovat jiné požadavky na autenticitu než na uchovávání a půjčování knih v knihovnách.<sup>65</sup>

Z tohoto hlediska je přesnější definice autenticity, kterou uvádí norma ISO 14721:2012 (a to jako jediný ze sledovaných zdrojů): „Míra, do níž osoba (nebo systém)

---

<sup>65</sup> Při koupi může například vědec očekávat, že knihu odborné literatury skutečně napsal uvedený autor (a je tedy autentickým dílem uvedeného autora, a nikoliv plagiátem); antikvariát může očekávat, že kniha nabízená prodávajícím je skutečně knihou, která byla kdysi vydána a nikoliv padělek.

pokládá objekt za to, za co je vydáván.“ (10, s. 20). To znamená, že v kontextu institucionalizovaného uchovávání musí být autenticita primárně otázkou toho, za co dokument vydává uchovávající organizace, například při zpřístupňování svým čtenářům (půjčování tištěných knih nebo prezentaci digitalizátů v digitální knihovně). V modelu projektu InterPARES je tento aspekt odpovědnosti archivu obsažen implicitně – ve fázi, kdy archiválii přijímá archiv od jejího původce, musí posoudit její autenticitu, a teprve poté ji uloží a uchovává tak, aby udržoval její autenticitu v průběhu času. V normě ISO 14271:2012 i v modelu InterPARES je pak kladen důraz na nutnost uchovávat nejen dokument, ale také přidružené doklady o jeho autenticitě, což je požadavek institucionální správy dokumentů.<sup>66</sup> Tento důraz na dokumentaci je také další podporou argumentu, že v oblasti institucionalizované správy je autenticita dokumentu v první řadě otázkou toho, za co tento dokument vydává sama organizace.

### **6.1.2 Kniha jako dokument**

V této disertační práci je pojmána institucionalizovaná správa komplexně, tedy nejen jako uchovávání a zpřístupňování dokumentů, ale i jako jejich vytváření (digitalizace knih, které se nacházejí v novodobých fondech dané knihovny). Je proto vhodné nejprve krátce pojednat i o otázce autenticity tištěných knih, které získávají knihovny.

Otázka autenticity fyzických dokumentů získávaných knihovnami podle uvedených studií nepředstavovala pro paměťové instituce vážnější problém (81) (86) vzhledem k technologické obtížnosti jejich zfalšování a osvědčeným postupům získávání dokumentů z ověřených zdrojů. Tyto studie mají na mysli zejména starší dokumenty (rukopisy apod.); pro tištěné knihy je toto riziko dále sníženo motivací takového činu (řádově nižší cena ve srovnání s rukopisy), existencí více exemplářů téhož vydání nebo zavedeným institutem povinného výtisku, v rámci něž knihovny získávají knihy přímo od vydavatelů.

I na uchovávání tištěných knih lze (v obecnější rovině) aplikovat požadavky archivu OAIS. Knihovny tyto požadavky naplňují například vytvářením metadat zachycujících identifikační informace (bibliografické záznamy) a kontextuální informace (systémy autorit). Otázky autenticity tištěných knih jsou v knihovnách především otázkou autenticity knihy jakožto datového objektu (exempláře vydání) a nezahrnují například otázky původce díla. Kniha musí být uchovávána v podobě, v jaké byla vydána, jako součást kulturního

---

<sup>66</sup> Od objektů uchovávaných soukromými osobami pro svoje potřeby zpravidla neočekáváme, že budou spojeny s dokumentací.

dědictví (bez ohledu na to, že jde o plagiát), a tedy jsou postačující osvědčené postupy pro získání knih (např. přímo od vydavatelů). To obecně platí zejména pro knihovny s povinným výtiskem, i když bychom mohli říci, že existují určité výjimky (např. systém jmenných autorit může uvádět skutečné jméno autora, který je v knize uveden pod pseudonymem). Pro některé jiné knihovny (např. vědecké knihovny) mohou existovat přísnější kritéria autenticity, zahrnující i otázku původce díla (uplatnění těchto kritérií pak může znamenat, že nebude pořízena kniha, o níž je známo, že obsahuje plagiát, nebo bude tato kniha vyřazena z fondu).

Otázka autenticity tištěných knih je tak především otázkou toho, za co je vydává konkrétní organizace prostřednictvím svého katalogu. Hlavním rizikem je tak chybná identifikace. Ta může v některých případech uvést čtenáře v omyl, který nemusí být při vypůjčení knihy na první pohled zřejmý (například jiné vydání nebo odlišné dílo se stejným názvem).

Z hlediska digitalizace však platí, že knihovny, které vytvářejí digitalizáty z knih svých novodobých fondů, jsou zodpovědné za řízení celého cyklu digitalizovaného dokumentu. To zahrnuje i otázku samotného vytváření tohoto dokumentu. Knihovny jsou tak ve fázi vytváření de facto v pozici vydavatelů, a tak otázky autenticity musejí být zohledněny již v této fázi, zejména v souvislosti s konceptem věrnosti.

Při revizi tematizací autenticity je potřeba vzít potaz rovněž odlišnost typů dokumentů, jimiž jsou knihy a archiválie. Následující popis obsahuje celou řadu zjednodušení pro zdůraznění podstatné odlišnosti těchto dvou typů dokumentů z hlediska jejich užití. Archivní dokumenty se od jiných typů dokumentů odlišují nikoliv svými formálními nebo obsahovými vlastnostmi, ale povahou svého vztahu k činnosti původce. Ve fázi, kdy jsou užívány v organizaci původce, slouží jako nástroj jeho činnosti. V okamžiku, kdy jsou vybrány k dlouhodobému uchovávání, stávají se součástí kulturního dědictví a jejich účel se mění, nicméně v obou případech spočívá hlavní význam archivního dokumentu v tom, že je dokladem určité skutečnosti; důležitější než tento dokument je skutečnost, o které vypovídá. Z tohoto hlediska lze archiválie označit za neautonomní dokument. To je také patrné na aktu převodu archiválií k dlouhodobému uchovávání, kdy je archiváři uspořádávají do smysluplných celků (sbírek) a opatřují archivním popisem, zahrnujícím informace o jejich původu (takto opatřenou archivní sbírku jako celek lze považovat za autonomní dokument). Archivní popis svou komplexitou výrazně přesahuje požadavky zpracování knih při katalogizaci. Sdělení, za účelem jehož přenosu archivní dokument vznikl, je obvykle známo jen omezenému okruhu osob, resp. není dostatečně srozumitelné bez znalosti kontextu

původního užití (který je znám jen původci, případně později v optimální situaci zaznamenán v archivním popise), archivní dokument obvykle existuje jen v několika málo exemplářích.

Institut vydávání knih znamená, že tatáž kniha je vydávána ve větším počtu identických exemplářů, a totéž dílo může být reprezentováno v průběhu času v různých vydáních. Knihy jsou reprezentací duševního díla, což o archiváliích většinou neplatí.<sup>67</sup> Kniha tak představuje relativně autonomní objekt; dílo, které reprezentuje, je srozumitelné samo o sobě širokému okruhu osob, a s jeho obsahem bývá také seznámeno mnoho čtenářů.

## **6.2 Revize tematizací autenticity**

### **6.2.1 Model pro předmět autenticity**

Skutečnost, že je zapotřebí nejprve přesně definovat předmět autenticity, a teprve poté na základě tohoto zkoumat otázky jeho autenticity, ukazují uvedené archivářské projekty. Oba archivářské projekty vytvořily vlastní model archiválie. Ačkoliv tyto projekty vycházejí ze stejné tradice (klasická diplomatika), jejich modely se liší. Model dokumentu projektu InterPARES zahrnuje tři konstitutivní součásti: dokumentová forma, anotace a kontext. Tento model je určen pro specifický model dokumentu (archiválie) a vymezuje nezbytné součásti, které musí mít každá archiválie. Tyto součásti jsou svázány s pravidly původce nebo legislativními požadavky na archiválie.

Norma ISO 14721, jak bylo uvedeno již v Kapitole 2, obsahuje vlastní komplexní informační model. Co nemusí být na první pohled zřejmé, je skutečnost, že v rámci tohoto modelu je klíčovým objektem z hlediska autenticity informační obsah. Požadavky na autenticitu model OAIS vztahuje na všechny typy informačních objektů uchovávaných v archivu, ale jádrem uchovávání je informační obsah, z čehož lze odvodit, že jádrem uchovávání autenticity je autenticita informačního obsahu. Pro fyzické i digitální objekty platí, že informační obsah musí být uchováván současně s archivačními informacemi, z nichž konkrétně provenienční informace tvoří podle modelu OAIS součást dokladu o autenticitě.

Tištěné knihy v kontextu uchovávání lze označit za typ fyzického objektem CDO. Pro národní knihovny a další typy knihoven, které zpřístupňují tištěné knihy širokému okruhu čtenářů, je praxí uchovávání tištěné knihy jako fyzického objektu CDO, bez uchovávání interpretačních informací. Cílovou komunitou může být jakýkoliv občan země a otázka

---

<sup>67</sup> Jistě existuje řada výjimek, například diplomová práce.

srozumitelnosti informací reprezentovaných tištěnými knihami předpokládá z hlediska znalostní základny pouze znalost jazyka dané země. Vše ostatní není předmětem požadavků na uchovávání kladených na tyto knihovny, ale jiných disciplín. Knihovny však uchovávají archivační informace vztahující se k těmto tištěným knihám, zejména identifikační informace (bibliografický popis) a kontextuální informace (např. systém autorit a předmětová hesla).

Pro všechny digitální objekty CDO platí, že informačním obsahem model OAIS rozumí uložený informační obsah (tj. objekt CDO a metadata obsahující interpretační informace). Jejich interpretace nutně zahrnuje zpracování počítačovou technologií, zatímco fyzické objekty jsou přímo přístupné lidským smyslům. Z modelu OAIS však lze vyvodit, že pro některé typy digitálních objektů CDO je třeba odlišovat mezi uloženým informačním obsahem a reprodukováným informačním obsahem. Tak je tomu i v případě digitalizátů knih, kdy reprodukce obvykle znamená zobrazení digitalizované knihy v její struktuře s možností navigace mezi jednotlivými částmi. Existují však typy digitálních objektů, které nejsou určeny pro reprodukci lidským uživatelům, ale pouze pro další zpracování jinými systémy. Mezi ně patří datové sady vědeckých archivů obsahující obrovské množství údajů zaznamenaných z pozorování. Norma ISO 14721 rozlišení na reprodukované a nereprodukované objekty neuvádí, ale upozorňuje na ně její hlavní autor ve své výkladové publikaci (80, s. 33-37).

V modelu standardu PREMIS pojem „intelektuální entita“ odpovídá reprodukovanému informačnímu obsahu, „reprezentace“ objektu CDO a „soubor“ částem, ze kterých se objekt CDO skládá. Elementy tohoto standardu jsou pak určeny jednak pro záznam interpretačních informací, jednak archivačních informací. Z hlediska interpretačních informací elementy standardu PREMIS zaznamenávají jen obecné informace, které lze uplatnit při popisu jakéhokoliv typu digitálního objektu. Pro zápis interpretačních informací o specifických typech těchto objektů slouží v kontextu současné praxe digitalizace zejména elementy standardu MIX.

V modelu InterPARES je jedním z typů kontextu archiválie provenienční kontext. Informační model OAIS uvádí, že provenienční informace lze chápat jako specifický typ kontextuálních informací, což ukazuje, že inspirací modelu OAIS je model projektu InterPARES. V modelu standardu PREMIS je navíc intelektuální entita chápána jako smysluplná jednotka intelektuálního obsahu (tedy dokument), což znamená specifikaci informačního obsahu modelu OAIS. Intelektuální entitu tak lze chápat jako podtyp informačního obsahu.



### 6.2.2 Časoprostorový aspekt

Základními prvky autenticity, které jsou explicitně (nebo někdy implicitně) obsaženy ve všech tematizacích, je otázka času (uchovávání) a prostoru (zpřístupňování). Časoprostorový aspekt je však explicitně a komplexně pojednán pouze v modelu InterPARES a normě ISO 14721:2012. Model InterPARES upozorňuje, že autenticita je ohrožena, kdykoliv dochází k přenosu dokumentů prostorem nebo v čase.

Funkční model OAIS popisuje dostatečně podrobný model pro zachycení tohoto časoprostorového vztahu. Časový aspekt je obsažen v posloupnosti vytváření balíčků SIP, AIP a DIP. Archiv vytváří z balíčku SIP, který mu dodá producent, balíček AIP první verze a ukládá je do funkčního celku Archivní úložiště. Balíček AIP se následně přesouvá do dočasného úložiště funkčního celku Zpřístupnění, kde se z něj vytváří balíček DIP, který se zpřístupňuje koncovým uživatelům. V průběhu těchto převodů balíčků může dojít k formátové konverzi objektu CDO, buď normalizací při převodu balíčku SIP do balíčku AIP (tj. ve funkčním celku Příjem konverzí formátu objektu CDO v balíčku IP do objektu CDO v archivačním formátu balíčku AIP), nebo vytvořením prezentačních formátů (tj. ve funkčním celku Zpřístupnění formátovou konverzí archivačního formátu objektu CDO balíčku AIP do prezentačních formátů objektu CDO v balíčku DIP).

V případě, že je nutné z hlediska uchovávání provést transformaci informačního obsahu balíčku AIP (např. formátovou konverzi objektu CDO ze zastarávajícího formátu do nového archivačního formátu), vzniká nová verze balíčku AIP. Z ní by měl být vytvořen nový balíček DIP. Nový balíček DIP může být vytvářen také v případě, že požadavky cílové komunity se změní (např. prezentační formát objektu CDO balíčku DIP formát zastará). V případě digitálních knihoven je balíček DIP načten jejich systémem a prezentován čtenářům v podobě zobrazených knih.

Prostorový aspekt má v modelu OAIS dvě podoby. První podobou je vztah mezi balíčkem AIP a balíčkem DIP. Tento vztah lze dále dělit na vztah mezi uloženým balíčkem AIP a uloženým balíčkem DIP nebo uloženým balíčkem AIP a zobrazeným informačním obsahem balíčku DIP nebo uloženým balíčkem DIP a zobrazeným informačním obsahem balíčku DIP. V případě digitálních knihoven je stěžejním zobrazený informační obsah balíčku DIP, který vidí čtenáři.

Druhou podobou je vztah mezi balíčkem AIP a jeho zálohami na fyzicky oddělených zařízeních. Zde může být otázkou, jaká kopie je směrodatná. Tento problém je zmíněn v modelu projektu InterPARES (v souvislosti s legislativními potřebami). V současné praxi

knihoven nejde o směrodatnost z hlediska právního systému, ale řešení případů, z jaké kopie se např. obnovují balíčky AIP po havárii hlavního úložiště. Těmto otázkám se však v této práci věnovat nebudeme. Jsou dostatečně odborně pojednány v literatuře.

Na tomto místě je také vhodné poznamenat, že prostorový vztah mezi současně existujícími objekty platí i pro tištěné knihy (exempláře téhož vydání uložené v různých knihovnách) a řadu dalších typů datových objektů, které jsou sériově vyráběny.

### 6.2.3 Otázka originálu

Otázka originálu je známou a nesčetněkrát tematizovaným aspektem autenticity fyzických objektů, které jsou lidskými výtvoři. Musíme odlišovat dva typy originálů. Prvním je originál ve smyslu jedinečného originálu, tedy fyzického objektu, který byl vytvořen v jediném „exempláři“. Do této oblasti spadají výtvarná díla a některé další typy objektů (např. rukopisy). Tato oblast je jádrem Benjaminovy konceptualizace nereprodukovatelného jedinečného originálu, který je konkrétně časoprostorově situován do konkrétního historického kontextu, ve kterém vznikl a jehož je odrazem. Druhým typem jsou sériově produkováné fyzické objekty, kam spadají i tištěné knihy. Tištěné knihy téhož vydání (tedy všechny jednotlivé exempláře) jsou považovány za rovnocenné fyzické originály v době svého vzniku, resp. distribuce.

V oblasti uchovávání fyzických dokumentů je problémem jejich nevyhnutelná degradace v průběhu času, kterou lze pouze zpomalit, nikoliv zastavit. Archivy proto přistupují k technice vytváření kopií těchto objektů. Fyzické originály jsou uchovávány v přísně kontrolovaných podmínkách a uživatelům jsou zpřístupňovány (např. půjčovány nebo prezentovány v galerii) pouze fyzické kopie. Některé knihovny dokonce přistupují k tomu, že rozpadající se noviny skartují a uchovávají pouze jejich surogáty v podobě mikrofilmových kopií.<sup>68</sup> Digitalizované dokumenty byly původně zvažovány jako možný surogát, který přetrvá poté, co se rozpadnou knihy a noviny vytištěné na kyselém papíře, ale tento přístup již není uznáván, vzhledem k později zjištěným problémům udržitelnosti digitálních dokumentů.<sup>69</sup> Tyto problémy vedly ostatně ke vzniku oboru digitální archivace.

Důležitým aspektem fyzického objektu je však také časové hledisko. Současná podoba fyzického originálu se liší od podoby, kterou měl, když byl vytvořen. Sto let stará kniha vypadá jinak než v době, kdy byla prodávána čtenářům. U knih lze dále odlišit okamžik

---

<sup>68</sup> Jak uvádějí například Deeganová a Tanner (81).

<sup>69</sup> Srv. (1, s. 32-33)

vzniku (např. v tiskárně) a okamžik, od kterého je nabízena k prodeji v knihkupectvích. Změny mezi těmito okamžiky nemusejí být vůbec patrné, ale také mohou (například vinou špatného zacházení při distribuci knih). Pro oblast institucionalizované správy je klíčovým okamžikem zařazení tištěné knihy do fondu knihovny. V této fázi je však také exemplář „obohacen“ o prvky dodané z důvodu potřeb správy (signatura, čárový kód, razítko knihovny apod.), takže jednotlivé knihovny uchovávají přísně vzato odlišné originály. Obecně se předpokládá, že čtenáři od těchto přidaných prvků odhlížíjí a nepovažují je za prvky originálu. To ukazuje na skutečnost, že znalostní základna čtenářů je dostatečná na to, aby byli schopni jako informační obsah přijímat pouze to, co tvoří původní prvky knihy. Postupem času se však tištěná kniha nutně mění od své podoby, ve které byla zařazena do fondu. V tomto smyslu je užitečné převzít Gladneyho koncept současného stavu fyzického objektu jako kopie originálu, kterým byl v době svého vzniku.

Tematizace autenticity digitálních dokumentů tak musí nutně řešit problém, jak s konceptem originálu naložit v kontextu digitálních technologií. Zde je možné vytvářet dokonalé kopie digitálních dokumentů, které se od sebe bitově nijak neliší (a tedy jedinečný originál zde nemůže existovat).

V normě ISO 14721 je za originál považován balíček AIP první verze (10, s. 105). Z toho vyplývá, že informační obsah v něm uložený (objekt CDO a metadata obsahující jeho interpretační informace) je originálem. Určitým problémem může být, že při vytváření balíčku AIP z balíčku SIP, vytvořeného producentem, může dojít k formátové normalizaci objektů CDO do archivačního formátu. V případě, že producent a archiv nejsou jedna a tatáž organizace a producent data nevytváří v archivačním formátu, je otázka této normalizace primárně otázkou vyjednání dohody o dodávání dat, která postup normalizace popisuje (nebo odkazuje na postupy uvedené v pravidlech uchovávání archivu), a vzájemné důvěry v nastavený konverzní proces. V současné praxi digitalizace, kdy je jedna a tatáž organizace zodpovědná za celý životní cyklus a v produkci se vytvářejí finální produkční data v archivačním formátu, tato otázka není relevantní. Pokud je objekt CDO v balíčku AIP první verze označován za originál, pak okamžikem pro určení časového aspektu originálu je datum a čas uložení balíčku AIP do funkčního celku Archivní úložiště. Vzhledem k tomu, že bitová ochrana archivu zahrnuje vytváření bitově identických záložních kopií balíčků AIP, musejí postupy archivu stanovit, co je směrodatnou kopií (případně může být více směrodatných), a to například z hlediska potřeby obnovy archivu po havárii.

Model InterPARES označuje za digitální originál takový dokument, který dosáhl svého účelu v rámci jeho užití při činnosti původce. Ten je nazván jako „první úplný a účinný dokument“ (76, s. 5). Z tohoto důvodu archiv při přebírání dokumentu od původce pracuje již pouze s kopií tohoto dokumentu. Výzkumníci projektu InterPARES si skutečnost, že archiv má pouze kopie dokumentu, zjevně obtížně zvykali vzhledem k tomu, že ve většině textu hovoří o kopiích dokumentu, ale v samém závěru (dokonce v poznámce pod čarou) nakonec uvádějí: „Každé kopírování, které následuje po akvizici, sice vytváří kopii archiválie, ale vzhledem k tomu, že kopie je tím jediným, co existuje, měla by se běžně označovat za archiválii, nikoliv za kopii“ (79, s. 11).

V případě digitalizovaných dokumentů se otázka originálu komplikuje v tom smyslu, že na jedné straně je zde fyzická předloha, která je převáděna, na druhé straně výsledný digitalizát knihy. Fyzická předloha se v současné praxi digitalizuje ve stavu, v jakém se nachází v době digitalizace (s případnými úpravami, například vyčištěním od prachu), a tato podoba má na výsledný digitalizát vliv (tato podoba je snímána). Pokud bychom aplikovali koncept digitálního originálu modelu InterPARES (originál jako první úplný a účinný elektronický archivní dokument), odpovídal by mu informační obsah, který je obsažen v balíčku SIP. V této podobě již prošel všemi úpravami a může být vydáván za hotový digitalizát (úplnost), který má schopnost (po reprodukci) digitálně reprezentovat tištěnou knihu (účinnost). Taková aplikace se zdá být bezproblémová.

Výše uvedené teze však ukazují, že je vhodnější pracovat s termínem „referenční objekt“. Konceptu referenčního objektu částečně odpovídá Rothenbergův koncept digitálního originálu (jako souboru vlastností, které musí sdílet všechny jeho autentické reprezentace) a Gladneyho koncept aktuálního fyzického objektu jako kopie objektu z doby jeho vzniku (tedy konceptuální objekt). Pro fyzické objekty nestačí určit samotný objekt, který je originálem, ale též časový aspekt. Totéž platí pro digitální objekty, pro které se však situace komplikuje tím, že je nelze zachovat v současném formátu, a v případě digitálních objektů, které vznikly digitalizací, zde vždy hraje roli také otázka jejich fyzické předlohy.

Z výše uvedeného a z hlediska konceptu věrnosti popsaného v tematizacích autenticity v Kapitole 5 je zřejmé, že referenční objekt jakéhokoli digitalizátu je komplexní objekt skládající se z několika odlišných složek. Tyto složky jsou: fyzická předloha v určený okamžik času (jako zdroj); informační obsah, který je cílem digitalizace a následného dlouhodobého uchovávání, a který je uložen balíčku SIP; popisu vztahu digitalizátu (informačního obsahu uloženého v balíčku SIP) k tištěné předloze. Uvedené vymezení platí

v případě, že digitální objekt CDO v balíčku AIP je v archivačním formátu, a tedy archiv neprovádí jeho formátovou normalizaci.

Pro digitalizáty knih je situace ještě komplexnější v tom smyslu, že tištěnou knihu lze pojímat v několika úrovních abstrakce, a podle toho řídit věrnost digitalizačního převodu (tyto úrovně abstrakce jsou popsány modelem FRBR). Popis vztahu digitalizátu k předloze pak znamená určení úrovně abstrakce, která má být cílem digitalizace, a způsobu odvozování za účelem dosažení tohoto cíle. Popis tohoto vztahu je tak základem pro posuzování věrnosti předloze (viz oddíl 6.2.5). Bližší specifikaci otázek referenčního objektu a věrnosti pro digitalizáty knihy se pak věnujeme v Kapitole 7.

#### 6.2.4 Aspekt odvozeniny

Časoprostorový vztah mezi aktuálním objektem a referenčním objektem je základním rámcem pro popis autenticity. Aktuální objekt je v tomto rámci vždy odvozeninou referenčního objektu, což platí i pro fyzické objekty (viz výše Gladneyho koncept aktuálního stavu fyzického objektu jako kopie objektu v době jeho vzniku). Požadavky na autenticitu odvozeniny znamenají posouzení autenticity odvozeniny vůči jejímu referenčnímu objektu a udržování autenticity v průběhu procesu odvozování od referenčního objektu do aktuálního objektu (řetězec odvozování). Pokud by mezi referenčním a aktuálním objektem neexistovala významná distance (časová, prostorová nebo časoprostorová), otázka autenticity postrádala smysl.

Standard PREMIS odlišuje dva typy digitálních odvozenin: odvozeniny vzniklé replikací (*replication*), které jsou bitově identické se zdrojovým objektem, a odvozeniny vzniklé transformací (*transformation*), které jsou bitově odlišné od zdrojového objektu. Archivační opatření založené na replikaci označuje standard PREMIS za bitovou ochranu (*bit preservation*), archivační opatření založené na transformaci za logickou ochranu (*logical preservation*). Norma ISO 14721 uvádí aspekt odvozování v souvislosti s celým balíčkem AIP, jehož odvozování archiv OAIS vykonává jako archivační opatření. Všechna tato odvozování řadí pod kategorii migrace (*migration*). Výraz „replikace“ standardu PREMIS tak odpovídá pojmům „renovace“ (*refreshing*) a „replikace“ (*replication*) normy ISO 14721, výraz „transformace“ (*transformation*) je užit ve shodném významu. V prvním archivářském projektu je užit výraz „replikace“ (*copying*) ve shodě s užitím v modelu PREMIS a pojem konverze (*migration*) ve shodě s užitím výrazu „transformace“.

V této práci budeme nadále užívat pojem replikace (tedy odhlížíme od rozdílů, které stanovuje norma ISO 14721, ale které jsou nepodstatné pro účely této disertační práce)

ve významu užitém ve standardu PREMIS, pojmy bitová a logická ochrana ve významu uváděném výše dle standardu PREMIS. Pojem replikace však užíváme pro odvozování v rámci celého životního cyklu digitalizátů knih (tedy i mimo kontext archivačních opatření v archivu OAIS), rovněž tak pojem transformace, který navíc užíváme v širším smyslu, jako transformace jakéhokoliv digitálního objektu v jakékoliv fázi životního cyklu. Typem transformace je formátová konverze (převod z jednoho formátu do druhého) vykonávaná v archivu nebo při digitalizaci (nebo při vytváření prezentačních variant v archivu). Také digitalizace je formou transformace (převod fyzického objektu do digitálního).

V digitálním prostředí obecně existují standardizované postupy odvozování, ať v případě replikace, nebo transformace. Pokud by neexistovaly, nebyla by mezi uživateli žádná důvěra v digitální technologie. Úspěch replikace očekáváme vždy, když stahujeme digitální data z internetu nebo je kopírujeme z jednoho počítače na druhý. Nástroje pro formátovou konverzi jsou navrhovány tak, aby přinášely užitek pro běžné užití, a obsahují vždy určitou racionalitu (jiné nástroje pro konverzi obrazových dat, jiné pro zvukové nahrávky apod.).

Požadavky paměťových institucí na uchovávání znamenají zavedení vyššího stupně zabezpečení do již existujících fungujících běžných postupů a vyšších nároků na procesy odvozování, které jsou odborně stanoveny s ohledem na dlouhodobý horizont. V případě replikace jednotně určují potřebu užití mechanismů hašovací funkce nebo využití vhodných datových nosičů a jejich včasnou výměnu. V případě formátových konverzí či jiných transformací je však úskalím digitální archivace je stanovit tak, aby přinášely užitek z dlouhodobého uchovávání. O tomto druhém problému pojednává oddíl 6.2.6.

Specifickým problém jsou otázky digitalizace, které znamenají převod vlastností fyzického objektu do objektu, tedy digitalizace knih. Zde jde o otázku věrnosti digitalizátu vůči předloze a také o skutečnost, že digitalizací se v současné praxi přidávají vlastnosti, které tištěná kniha neměla (např. plnotextová prohledatelnost). O těchto oblastech pojednávají následující dva oddíly.

### **6.2.5 Koncept věrnosti digitalizace**

Otázka věrnosti (ve smyslu vztahu digitálního objektu CDO a tištěné knihy) vyvstává přísně vzato nejdříve v okamžiku, kdy jsou vytvořena a zkontrolována finální produkční data; je tedy otázkou informačního obsahu v balíčku SIP. Prvním krokem pro toto posuzování však musí být stanovení, na jaké úrovni abstrakce je tištěná kniha snímána. Uvedené tematizace věrnosti se implicitně vztahují k úrovním exempláře a vydání (blíže viz Kapitola 7).

Uvedené tematizace věrnosti lze shrnout do dvou oblastí: otázka věrnosti reprodukování informačního obsahu (to, co je vnímatelné uživatelem) a otázka, jakým způsobem je věrnost digitálně specifikována (tj. na úrovni uloženého informačního objektu, který se skládá z objektu CDO a metadat obsahujících interpretační informace).

První oblast se týká převáděných vlastností tištěné knihy. Je zřejmé, že digitalizací lze přenést pouze obrazové a strukturální vlastnosti. Uvedené tematizace se shodují v tom, že je nutno zachovat barevnou věrnost předloze, dostatečnou čitelnost písma, původní a úplnou posloupnost stran tištěné knihy (tj. i pokud je paginace v knize chybná a včetně vakátů a přídeští) a úplnost stran včetně zachování malé části za hranou stránky, ze které lze odlišit okraj stránky (tj. neprovádět ořezy).

Grycz uvádí specifické pojetí strukturálního modelu reprezentace, kdy základním strukturním prvkem je snímek dvojstránky. Podle Grycze lze pouze tímto způsobem autenticky reprezentovat fyzickou předlohu, tj. v podobě, v jaké ji běžně zakoušejí její čtenáři. Tento způsob není v současné praxi běžný. Pokud lze toto považovat za otázku nižší věrnosti, pak částečným řešením může být opětovné spojení samostatně uložených snímků do dvojstránky při zobrazení v digitální knihovně, které je technicky snadno proveditelné.<sup>70</sup> Na jinou stranu problému věrnosti odkazuje destruktivní digitalizace, která se vyskytuje v některých digitalizacích současné praxe. Destruktivní digitalizace umožňuje zachytit stránku v její úplnosti (tj. i v částech, kde je spojení stránek s vazbou). Problémem této metody je, že ji nelze aplikovat vždy.

Druhá oblast souvisí s možnostmi a omezeními technické specifikace digitální reprezentace (tedy uloženého informačního obsahu). Bitová hloubka a barevný model nepředstavují větší problém. Nejen směrnice FADGI, ale řada dalších směrnic doporučují jako standard pro barevné snímky bitovou hloubku 24 bitů (8 bitů pro 3 barevné kanály) a barevný model RGB. Prostorové rozlišení musí být určeno na základě znalosti předlohy, takže postačuje konstatování, že doporučená minimální rozlišení pro různě definované skupiny fyzických předloh jsou nastavena paušálně a pro nestandardní fyzické dokumenty může být potřeba toto doporučení změnit. Specifickým problémem je otázka formátu a barevného profilu, který je potřeba v kontextu digitální archivace vybrat nejen z hlediska věrnosti předloze, ale také s ohledem na požadavky dlouhodobého uchovávání. V tomto ohledu se zdá být vhodnějším doporučení směrnice FADGI, aby rastrové obrazy výsledných

---

<sup>70</sup> Tento přístup je využit například v alternativním českém klientovi <http://www.digitalniknihovna.cz/>.

digitalizátů byly uloženy v barevném profilu Adobe RGB 1998. Tento profil je široce užívaný a podporovaný. Zde je třeba dodat, že tento ICC profil pro barevný prostor Adobe RGB má svoji oficiální podobu, kterou specifikovala firma Adobe a která nese název „Adobe RGB (1998) ICC profile“ (91, s. 2). Jiné subjekty sice mohou vytvořit vlastní ICC profil pro namapování do barevného prostoru Adobe RGB 1998, ale nesmějí užívat tento název. Z hlediska důvěryhodnosti původce je tedy vhodné využít tento oficiální ICC profil, jehož specifikace je volně dostupná na webovém sídle firmy Adobe (91, s. 17). V oblasti formátů jsou doporučovanými archivačními formáty TIFF a JP2 (viz oddíl 3.1.3).

#### **6.2.6 Klíčové vlastnosti**

Typ odvozování, kterým je transformace, znamená bitovou změnu digitálního objektu. Jádro problémů digitální archivace spočívá v tom, že v průběhu času bude potřeba provést transformaci původního digitálního objektu CDO (který je obsažen v balíčku SIP, resp. AIP první verze), která mění původní bity, čímž vyvstává otázka, zda nebyl porušen původní informační obsah. Do této oblasti spadá zejména otázku formátových konverzí.

Souvislost mezi transformací a autenticitou uvádějí především tematizace normy OAIS, standardu PREMIS a projektu InterPARES v podobě konceptu vlastností, které mají být zachovány napříč transformacemi.

V novém vydání ISO 14721 jde o pojmy popis vlastností informací a transformační vlastnost informací. První znamená popis těch vlastností informací informačního objektu, které mají být zachovány v průběhu uchovávání, druhé specificky ty vlastnosti, které se stanovují jako cíl nevratné transformace. Ty pak podle normy ISO 14721:2012 mohou posloužit jako doklad o autenticitě. Stejný koncept je obsažen ve standardu PREMIS (signifikantní vlastnosti), který jej zavedl již dříve s tím rozdílem, že pojem transformační vlastnost informací reflektuje skutečnost, že některé jiné vlastnosti budou nutně ztraceny.

V modelu projektu InterPARES, který je ze sledovaných tematizací nejstarší, se uvádí, že podmínkou zachování integrity je, aby byla archiválie úplná a neporušená ve všech podstatných aspektech. Tomu odpovídá jeden z jeho požadavků pro udržování autenticity: zachovat obsah dokumentu, všechny vyžadované anotace a znaky dokumentové formy po skončení každého procesu odvozování (požadavek „B.1.c“ v oddíle 5.2.2.5). Ačkoliv výraz „podstatné aspekty“ (*essential aspects*) není definován v terminologickém slovníku projektu, z výše uvedeného vyplývá podobnost s konceptem signifikantních vlastností.



Problém je, jakým způsobem určit, které vlastnosti jsou důležité pro daný typ dokumentu, a tedy musejí být zachovány napříč transformacemi, a také to, jak tyto vlastnosti přesně definovat. Ve standardu PREMIS je možno tyto vlastnosti definovat pro všechny úrovně objektu (intelektuální entita, reprezentace, soubor, bitový tok). Elementy PREMIS jsou pouze dva: typ signifikantní vlastnosti a její hodnota. To nabízí naprostou volnost v jejich užití, ale současně neposkytuje žádné bližší vodítko. Standard pouze uvádí, že signifikantní vlastnosti lze definovat na různých úrovních abstrakce (např. typ: „obsah“, hodnota „veškerý textový obsah a obrázky“; nebo typ: „počet stran“ a hodnota „7“) (52, s. 51). Žádné externí metadatové schéma pro rozšíření schématu PREMIS (pro konkretizaci signifikantních vlastností určitého typu dokumentu) neexistuje. Stanovení signifikantních vlastností navíc PREMIS označuje za subjektivní.

Jakkoliv je koncept signifikantních vlastností složitý a v současnosti nedostatečně probádaný, je nepochybně užitečný přinejmenším k tomu, aby ukázal na nutnost před každou transformací explicitně stanovit cíle, kterých má být dosaženo, a na základě nich výsledky transformace kontrolovat. Pokud jsou klíčové vlastnosti relevantní, pak jejich nezachování lze jednoznačně považovat za narušení autenticity.

V této práci budeme dále užívat pojem klíčové vlastnosti pro označení vlastností, které mají být zachovány při transformaci. V rámci řízení celého životního cyklu digitalizátů je však potřeba vztáhnout tento pojem i na odvozování mimo archiv. Stanovení klíčových vlastností je důležité jako cíl digitalizace i jako cíl vytváření prezentačních variant. Při digitalizaci musí zahrnovat převáděné vlastnosti i vlastnosti, které jsou digitalizací vytvářeny.

### **6.2.7 Identita dokumentu**

Projekt InterPARES definuje identitu dokumentu jako celek atributů, které jej jedinečně charakterizují a odlišují od jiných dokumentů, přičemž identita je jedna ze dvou složek autenticity. Deeganová a Tanner považují za jedno z hlavních rizik pro autenticitu chybnou identifikaci dokumentů. Bearman a Trantová ve stejné souvislosti uvádějí problém existence různých verzí digitálních dokumentů nebo různých digitalizátů téže fyzické předlohy na internetu, které nejsou dostatečně odlišeny; řešení v tomto ohledu spatřují v perzistentních identifikátorech. S perzistentními identifikátory spojuje identitu také Caplanová. Gladney uvádí, že nejsme ochotni považovat za autentický takový dokument, u něhož není známa identita jeho tvůrce a konkrétní historická událost vzniku. Model OAIS

uvádí jako koncept identifikačních informací, které popisují informační obsah a které musejí být spolu s ním uchovávány v archivu.

V uvedených definicích je identita spojena s informačním obsahem. InterPARES přinesl výčet základních atributů, které musí mít každý dokument přijímaný do archivu; na jejich základě je posuzována identita. Tento výčet je specifický pro oblast archiválií, nicméně po určitém zobecnění lze v attributech archiválie spatřit paralelu s praxí bibliografického popisu. Bibliografický popis je popis atributů identity knih z hlediska jejich uchovávání.

Identita tedy souvisí s vlastností, že dokument je tím, za co je vydáván v dané organizaci, přičemž atributy identity musejí být explicitně vyjádřeny. V případě správy tištěných knih v knihovnách to znamená vyjádření bibliografickým záznamem v katalogu. Identitu jako celek charakterizujících a odlišujících atributů je však možno zvažovat i na úrovni digitálního objektu CDO – jako digitální identitu. Digitální identitu je v tomto případě potřeba rozdělit do dvou skupin: konkrétní digitální identita a typová digitální identita. Typová digitální identita znamená specifikaci určitého prvku, parametru, vlastnosti nebo funkce digitálního objektu, který lze aplikovat na více objektů (např. konkrétní formát). Konkrétní digitální identita je souhrn atributů konkrétního (na konkrétním nosiči uloženého) digitálního objektu, které jsou hodnotami typové digitální identity. Možnosti popisu konkrétní digitální identity vymezují metadatové standardy (například PREMIS nebo METS). Nad rámec těchto standardů lze vytvořit vlastní dílčí externí schémata, která mohou tyto atributy podrobněji popisovat. Digitální identitu bude v průběhu času nutné změnit (viz formátové konverze v archivu z důvodu zastarávání archivačního formátu v balíčku AIP), nicméně v daný okamžik určuje aktuální existenci digitálního objektu CDO, která je předpokladem existence informačního obsahu.

### **6.2.8 Integrita dokumentu**

Souvislost mezi integritou a autenticitou je v uvedených tematizacích pojata velmi různorodě. V prvním zmíněném archivářském projektu integrita archiválie znamená její spolehlivost a autenticitu. V projektu InterPARES tvoří integrita s identitou dvě základní složky autenticity. Ve standardu PREMIS znamená autenticita objektu možnost ověřit jeho integritu. V jiných tematizacích jsou autenticita a integrita pojímány jako dvě odlišné kategorie (např. ve směrnici DSA).<sup>71</sup>

---

<sup>71</sup> Pro přehled viz také (1, s. 57-58).

Integrita dokumentu je podle projektu InterPARES jeho úplnost a neporušenost ve všech podstatných ohledech, což odkazuje k výše uvedenému konceptu klíčových vlastností: napříč transformacemi se mění původní bity, ale musejí být zachovány stanovené klíčové vlastnosti.

Z koncepcí identity a integrity vyplývá, že posuzování integrity je založeno na znalosti identity objektu – nejprve musí být známo, o jaký objekt jde, a teprve poté lze posuzovat, zda nebyl porušen nebo zfalšován. Otázka autenticity objektu, o kterém nic nevíme, je tedy sama o sobě problematická. Pro oblast institucionalizovaného uchovávání lze tedy vyvodit, že organizace by měla vyvinout maximální úsilí při zjišťování (ve fázi akvizice), stanovování (ve fázi vlastní produkce) a popisu identity konkrétního dokumentu. Tento popis by měl zřejmě zahrnovat i stanovení klíčových vlastností.

Specifickým problémem, který se na základě analýzy v této kapitole stává dílčí výzkumnou otázkou této disertační práce, je zjistit, co tvoří identitu digitalizátu knihy (a to jak identitu informačního obsahu, tak digitální identitu) a jak definovat a popsat její nezbytné atributy. Knihovna, která je sama producentem digitalizátů, má situaci relativně jednodušší v tom, že jde o její digitální dokumenty, a proto může být operace stanovení klíčových vlastností méně kontroverzní než pro knihovny, které uchovávají dokumenty jiných organizací.

### **6.2.9 Koncept provenience**

V projektu InterPARES byl jako jedna z konstitutivních složek archiválie definován kontext. Za jeden z typů kontextu pak byl označen provenienční kontext, který tvoří původce záznamu, jeho mandát, struktura a funkce. Tento kontext i ostatní kontexty jsou podle projektu předmětem archivního popisu konkrétní archivní sbírky (někdy tříděné podle původce).

Pojem provenience se v dalších tematizacích vyskytuje s různými rozdíly, například:

- historie událostí vztahujících se k digitálnímu objektu (8, s. 8);
- původ objektu a historie jeho změn (30, s. 55);
- historie objektu (52, s. 2);
- informace o tvůrci a události vzniku a historii odvozování (90, s. 97).

Definice normy ISO 14721:2012 de facto zahrnuje vše výše uvedené (původ / zdroj informačního obsahu, všechny jeho změny a informace o všech správcích). Podle normy je

odpovědností archivu OAIS vytvářet a uchovávat provenienční informace od okamžiku příjmu balíčku SIP od vkladatele; informace z předchozí doby by však měl dodat vkladatel. Tento koncept odpovědnosti má analogii v projektu InterPARES s tím rozdílem, že v normě ISO 14721 jsou odpovědnosti explicitně vyjádřeny v dohodě o dodávání dat. Jak již bylo také uvedeno, koncept provenienčních a kontextuálních informací modelu OAIS je převzat z archivnictví.

#### **6.2.10 Dokumentace a metadata**

V oblasti institucionalizované správy je rozhodným prvkem požadavku autenticity (jenž je explicitně či implicitně obsažen ve většině uvedených tematizací) nutnost vytvářet a udržovat patřičná metadata a dokumentaci. Metadata lze rovněž považovat za součást dokumentace, ale ve většině tematizací se dokumentací míní popis v přirozeném jazyce, zatímco metadaty strukturovaná data v umělém jazyce, která mají také svoji vlastní dokumentaci (tj. popis metadatového standardu). V některých tematizacích se dokumentace a metadata označují za de facto jedinou formou dokladu o autenticitě digitálního objektu (84).

Základním rámcem pro metadata i dokumentaci jsou koncepty interpretačních a archivačních informací modelu OAIS. Je potřeba zopakovat, že metadata zaznamenávající tyto typy informací jsou podle tohoto modelu stále „jen“ digitální data, tj. je potřeba, aby k nim také existovaly interpretační informace, což vede k nekonečné rekurzivitě (kterou může ukončit pouze to, že na konci řetězce dalších interpretačních informací jsou takové interpretační informace, které jsou zaznamenány v podobě fyzického objektu, jako je tištěné dokumentace – takový objekt je již vnímatelný člověkem bez potřeby počítačového zprostředkování). Tento problém však není zásadní vzhledem k tomu, že metadata jsou v současné praxi zaznamenávána v textových formátech (typicky v XML), jejichž interpretace počítačovým prostředím se nepovažuje za problematickou ani z dlouhodobého hlediska. Existující užívané standardy jsou dostatečným základem pro současnou praxi digitalizátů knih, ačkoliv vyžadují správný způsob převzetí do aplikačních metadatových profilů a další vývoj (zejména v oblasti signifikantních vlastností schématu PREMIS).

Sama tato metadata musejí být též autentická. Za tímto účelem se lze řídit doporučením zvýšit redundanci metadat jejich uložením do balíčku AIP i do databáze.<sup>72</sup> Z hlediska uchovávání autenticity je také důležité, aby balíčky AIP byly nezávislé na konkrétních technologiích užívaných pro archiv OAIS. To znamená, aby byly

---

<sup>72</sup> Uložení metadat v databázi mimoto podporuje rychlé vyhledávání. Srv. (121, s. 7).

samopopisné (tj. obsahovaly objekt CDO i dostatečná metadata) a oddělitelné od aktuálně užívané technologie (tuto možnost lze otestovat např. postupem pro obnovu při havárii).

V požadavcích na udržování autenticity archiválií projektu InterPARES je uvedena nutnost vytváření a udržování dokumentace procesů odvozování a jejich důsledků, který zahrnuje i potřebu zpřístupňovat uživatelům informace o procesu, jehož důsledkem odvozenina nepřenáší v úplnosti a věrně všechny znaky identity a integrity. Dále je zde uveden požadavek archivního popisu sbírek, který musí zahrnovat popis všech kontextů dokumentů (o jejich původci, legislativních rámci apod.) a souhrnné informace o všech změnách dokumentů od doby jejich vytvoření původcem. Z hlediska normy ISO 16363 je důležitá dokumentace pravidel uchovávání (ve kterých by měl archiv popisovat své postupy), dokumentace jednotlivých postupů a záznamy o provedených opatřeních. Do posledně jmenované kategorie spadají jak protokoly vytvářené pracovníky archivu, tak záznamy aplikací.

Záznamy aplikací jsou další formou dokumentace, vytvářené samotnými aplikacemi. Patří mezi ně provozní logy a zejména prověřovací záznam (*audit trails*), která projekt InterPARES definuje jako „prostředky pro sledování všech interakcí s dokumenty v elektronickém systému takovým způsobem, aby mohl být každý přístup do systému okamžitě zdokumentován, a to pro účely ochrany před neautorizovanými úkony s dokumenty.“ (76, s. 1). Norma ISO 14721 chápe prověřovací záznam širěji: veškeré provenienční informace mohou podle ní posloužit jako prověřovací záznam (10, s. 74)

Celkově lze shrnout, že z hlediska autenticity patří mezi nejdůležitější informace (které je nutno zaznamenat do metadat a do dokumentace) údaje o kompletním řetězci odvozování a všech užitých postupech, údaje o attributech identity informačního obsahu, údaje o attributech digitální identity a popis klíčových vlastností.

#### **6.2.11 Problémy reprodukce**

V oddíle věnovaném aspektu odvozeniny jsme se zabývali odvozováním informačního obsahu jakožto uloženého objektu. Odlišnou otázkou je problém počítačové reprodukce informačního obsahu z aktuálně uložené odvozeniny do podoby informačního obsahu, který může vnímat člověk (v kontextu současné praxe digitalizace tedy zejména vidět). Caplanová ve své studii uvádí, že objekt může být autentickým, i když není možné jej reprodukovat. To implikuje přístup, že autenticita podle ní nesouvisí s reprodukováním. Tento přístup je v kontextu zkoumaných tematizací autenticity spíše výjimkou. Norma ISO 14721, projekt

InterPARES a většina dalších tematizací (zejména těch, které se zabývají otázkou zpřístupňování v digitální knihovně) explicitně nebo implicitně zahrnuje do požadavků autenticity otázku reprodukce. Otázka autenticity je v případě digitalizátů knih (a jiných reprodukováných objektů) v konečném důsledku vždy otázkou autenticity reprodukováného (zobrazeného) informačního obsahu.

Projekt InterPARES stanovuje jako hlavní úkol uchovávání zachovat reprodukovatelnost dokumentu. Podobně je tomu v normě ISO 14721 s tím rozdílem, že tato norma zahrnuje, jak bylo uvedeno, i otázky uchovávání typů informačního obsahu, které nejsou určeny k reprodukci uživatelům, ale zpracování jinými aplikacemi (např. datové sady). Norma ISO 14721 v souvislosti s reprodukovánými objekty varuje před tím, aby se archiv OAIS nespolehal pouze na zpřístupňovací software jako na způsob řešení problému interpretačních informací. V případě zastarání tohoto softwaru se informační obsah nereprodukovatelným, a bude tedy nenávratně ztracen. V současné praxi digitalizace je však hlavním způsobem zpřístupňování prezentace v digitální knihovně, kdy užití zpřístupňovacího softwaru je nezbytnou součástí řízení životního cyklu digitalizátů knih.

Problémem v oblasti reprodukce informačního obsahu však není jen zastarání zpřístupňovacího softwaru, ale také skutečnost, že různé softwarové aplikace mohou tentýž objekt CDO interpretovat různě. Žádná z uvedených tematizací se nezabývá tímto problémem (tj. problémem autentického zobrazení). V praxi přitom nejde o výjimečný jev. Vyskytuje se například u reprodukce formátu PDF<sup>73</sup> nebo rastrových formátů.<sup>74</sup> Do určité míry jde o otázku kvality softwaru. Částečně může jít také o problém kvality dokumentace (tj. zda dostatečně detailně a jednoznačně popisuje daný formát). V případě komplexních formátů pak může být otázkou sama možnost provést jednoznačný popis, který by vylučoval odlišné volby při vytváření softwaru pro práci s daným formátem.

#### **6.2.12 Rizika pro udržení autenticity**

Jedním ze způsobů, jak zkoumat požadavky autenticity, je určit rizika, kterým je udržení autenticity vystaveno. Škála rizik popsáných v uvedených tematizacích je široká, některá jsou obecně známá, jiná nejsou samozřejmá. Zdrojem těchto rizik je digitální prostředí, ve kterém je snadnost kopírování, pozměňování nebo porušení digitálního dokumentu nesrovnatelně vyšší než ve sféře fyzických dokumentů.

---

<sup>73</sup> Srv. např. (114, s. 3-5)

<sup>74</sup> Srv. zprávu o testování zobrazení formátu JP2 v různých aplikacích (122, s. 98-108)

Rizika pro udržení autenticity, která uvádějí zkoumané tematizace, můžeme rozdělit do těchto kategorií:

- Chyby při kopírování na jiný nosič (vč. přenosu po síti)
- Nezáměrná chyba
- Záměna dokumentu
- Zfalšování
- Absence dokumentace změn
- Formátová konverze
- Chybná nebo nekvalitní identifikace
- Nejasný původ dokumentu
- Nekvalitní digitalizace

Zfalšování (tematizované tradičně v oblasti archivnictví) není hlavním rizikem pro autenticitu digitalizátů knih; záměna dokumentu (riziko založené na stejném principu) však ano. Formátová konverze může znamenat ztrátu důležitých vlastností, pokud cílový formát nedokáže zachovat klíčové vlastnosti původního formátu. Formátová konverze představuje také riziko pravděpodobnosti softwarové chyby, která nemusí být na první pohled patrná a může se odhalit (nebo projevit) až za několik let, kdy může být náprava mnohem obtížnější, než by bývala byla v době konverze. Živá debata o volbě základního přístupu k archivaci (migrace versus emulace) probíhala v mezinárodní komunitě v 90. letech 20. století a riziko chyb formátové konverze bylo jedním z hlavních argumentů ve prospěch emulace.

Specifickým případem rizik je chybná identifikace digitálního dokumentu, zejména vzhledem k tomu, že existují různé verze, které od sebe nejde odlišit, dokumenty jsou zpřístupňovány na různých webových sídlech. Chybná identifikace také hrozí v případě vytváření různých digitalizátů téže knihy (různými institucemi nebo projekty v průběhu času). Absence vhodného perzistentního identifikátoru (resp. identifikačního systému) byla v tomto ohledu vždy považovaná za velký problém a potřeba perzistentního identifikátoru je zmiňována v několika tematizacích.

Nejasný původ odkazuje jednak na prezentaci v internetovém prostředí (důvěryhodnost webových sídel) a obecněji také na špatný popis prezentovaných dokumentů.

V případě citování je běžným jevem, že nás odkaz zavede přímo na dokument ve formátu PDF, ze kterého lze jen obtížně zjistit, kdo je jeho původcem, neboť je součástí širšího celku, ke kterému se z odkazu na soubor v PDF nedostaneme. Paradoxně je to i případ webového sídla projektu InterPARES.

Dále je zde také problém nestability internetových adres (URL adres) dokumentů, které se v praxi často mění. URL adresu lze považovat za součást konkrétní digitální identity, která je dočasná. To vede k souvisejícímu problému důvěryhodnosti citační praxe.

### **6.2.13 Shrnutí revize a dílčí závěr**

Uvedená revize ukázala, že jednotlivé tematizace jsou v základních aspektech vzájemně kompatibilní.<sup>75</sup> To je dáno do značné míry tím, že tematizace vycházejí z tradiční definice autenticity užívané v archivnictví, ačkoliv v nich tento zdroj není většinou explicitně uvedený.

Podstatné prvky konceptu autenticity můžeme shrnout následujícím způsobem. Autenticita je vždy vztahem mezi referenčním objektem a aktuální objektem, který je od referenčního odvozen a jehož autenticita je posuzována, přičemž mezi oběma objekty musí existovat významný rozdíl nebo časoprostorová distance. Všechny tematizace předpokládají určitý referenční objekt (nazývaný jako originál nebo chápaný jako původní objekt, který je na počátku řetězce odvozování). Termín referenční objekt byl zaveden jednak jako vyjádření toho, že jde o objekt, ke kterému se aktuální objekt má vztahovat, jednak jako zpřesnění pojmu originál, který zahrnuje nutný aspekt, a sice času (současný originál obrazu je jiný než originál v době svého vzniku, z hlediska stárnutí), a nakonec také z důvodu odlišné povahy fyzických a digitálních objektů. Referenční objekt se navíc může skládat z více složek (což je právě případ digitalizátů knih – odkazujeme se k tištěné předloze, ale také k vlastnostem vytvořeného digitálního objektu, např. prohladatelnosti, které mají být zachovány).

Vztah mezi referenčním objektem a aktuálním objektem je také vyjádřen ve vztahu mezi identitou a integritou, který je spíše implicitně obsažen v archivářské konceptualizaci. Při posuzování je nejprve nutno zjistit identitu dokumentu, a poté prokázat jeho integritu. Je tedy zřejmé, že v oblasti institucionální správy nelze hovořit o autenticitě dokumentu bez přidružených informací, které popisují identitu objektu. Organizace vydává dokument za to, co uvádí v popisu jeho identity. V praxi správy tištěných knih jde tedy o bibliografický

---

<sup>75</sup> Jistou výjimkou je koncept Caplanové, podle něhož autenticita nesouvisí s reprodukováním.



záznam v katalogu. V kontextu digitalizátů knih je otázkou, jaké atributy tvoří jejich identitu a jak je zapsat, a dále jak zachovat jejich integritu digitalizátů. Ukázalo se, že bude potřeba pojem identity, který je v archivnictví spojen pouze s reprodukováným dokumentem, rozšířit i na identitu digitálního objektu CDO (digitální identitu). Koncept identity fyzických objektů CDO v kontextu běžné správy fyzických objektů postrádá smysl aplikovat.<sup>76</sup> V digitálním světě však má své opodstatnění, jak nejzřetelněji ukazuje problém identifikace formátů (tedy přesného zjištění identity formátu).

Formálně nejvytříbenější definici autenticity představují výstupy projektu InterPARES a Gladneyho definice tvrzení o autenticitě. Využitím těchto definic a jejich úpravou na základě revize provedené v předcházejících oddílech této kapitoly lze definovat autenticitu dokumentu následujícím způsobem: Autenticita je vlastnost (aktuálního) dokumentu, který je takovou odvozeninou od stanoveného referenčního objektu, která zachovává jeho klíčové vlastnosti při jeho uložení i jeho reprodukci počítačovou technologií, a jehož identita a provenience jsou dostatečně zdokumentovány (včetně popisu referenčního objektu, jeho klíčových vlastností a řetězce odvozování a správy).

Domníváme se, že tato definice není v rozporu s uvedenými dvěma zdrojovými tematizacemi, ale vhodně je doplňuje a rozšiřuje a zahrnuje všechny podstatné aspekty, které byly zjištěny při revizi zkoumaných tematizací autenticity. Tato definice bere v potaz kontext institucionální správy (odtud plyne potřeba současně s uchováváním dokumentu vytvářet a uchovávat také jeho dokumentaci). Definice autenticity normy ISO 14721:2012 uvádí, že autenticita je otázkou míry, do které může být objekt považován za autentický. Tato definice zohledňuje hledisko posuzování autenticity. Moje definice nezohledňuje určitou perspektivu autenticity (posuzování versus udržování autenticity). Zdá se, že problém definice klíčových vlastností i otázka úplnosti dokumentace s sebou implicitně nesou otázku míry autenticity a pravděpodobných částečných ztrát v procesu formátových konverzí v průběhu uchovávání. Cílem této práce však není definice autenticity. Její vytvoření slouží jako příklad toho, že z hlediska definice není vážnější rozpor napříč různými tematizacemi. Rozdíly spočívají spíše v míře detailu a preciznosti definic. Problémem je podrobnější specifikace požadavků autenticity pro oblast digitalizátů knih.

Revize provedená v této podkapitole ukázala, že pro vytvoření obecného rámce pro zkoumání autenticity v životním cyklu digitalizátů knih lze v plné míře a beze změn převzít

---

<sup>76</sup> Koncept fyzické identity by znamenal např. provést forenzní analýzu materiálu knihy a výsledek zapsat do dokumentace. Tento postup se aplikuje jen ve specifických případech.

jako základ rámec pro autenticitu vymezený normou ISO 14721:2012 (pojem autenticity a její vztah s provenienčními informacemi) s užitím vybraných konceptů dalších tematizací. Z výstupů projektu InterPARES jde zejména o koncepty identity a integrity a ze standardu PREMIS intelektuální entity a odvození. Jako výsledek revize byly vytvořeny vlastní koncepty referenčního objektu (namísto originálu) a digitální identity a z důvodu sjednocení byl zaveden výraz „klíčové vlastnosti“ (pro vlastnosti informačního obsahu, které mají být zachovány napříč všem transformacím).

### **6.3 Návrh obecného rámce pro zkoumání autenticity digitalizátů knih**

Jako vhodný konceptuální základ pro vytvoření obecného rámce pro zkoumání autenticity v životním cyklu digitalizátů knih byla shledána norma ISO 14721:2012 se svým informačním modelem a definicí autenticity. Tento rámec je dále rozšířen a zpřesněn užitím vybraných pojmů projektu InterPARES a PREMIS a užitím vlastních konceptů referenčního objektu a digitální identity.

#### **6.3.1 Model pro předmět autenticity v celém životním cyklu**

Hlavním předmětem autenticity je podle modelu OAIS informační obsah. Ten je tvořen digitálním objektem CDO a interpretačními informacemi. Informační obsah budeme dále odlišovat jako uložený informační obsah a reprodukováný informační obsah. Uložený informační obsah se skládá z digitálního objektu CDO a uložených interpretačních informací. Interpretační informace mohou být uloženy ve dvou podobách: v podobě technických metadat (tj. digitálních objektů reprezentujících informace např. o formátu nebo obrazových vlastnostech objektu CDO) a v podobě dokumentace, která může být uložena v digitální podobě (např. ve formátu PDF) nebo fyzické podobě (vytištěná dokumentace). Reprodukováný informační obsah budeme nazývat v souladu se standardem PREMIS intelektuální entitou, což je vnímatelný jednotlivý intelektuální nebo umělecký výtvar – tím je v kontextu této dizertace zobrazený digitalizát knihy (intelektuální entita může být podle PREMIS verze 3 reprezentována i fyzickým objektem, tedy i tištěná kniha je intelektuální entita). V pasážích, kdy je potřeba explicitně odlišit tyto dva typy intelektuální entity, uvádíme pojem digitální intelektuální entita pro reprodukováný informační obsah (zobrazený digitalizát) a fyzická intelektuální entita pro vnímaný informační obsah tištěné knihy. V ostatních případech uvádíme pouze výraz intelektuální entita ve smyslu digitální intelektuální entita.

Jako rámec pro vyjádření časoprostorového vztahu je plně převzat model vztahů mezi balíčky SIP, AIP a DIP normy ISO 14721, který byl shrnut v oddíle 6.2.2. Jeden tento balíček vždy obsahuje jeden uložený informační obsah (tím odlišujeme případy, kdy může být jednotlivý informační obsah dodán např. ve více balíčcích SIP, což norma ISO 14721 umožňuje). Specifikem této normy je, že informační obsah uložený v balíčku SIP jde nejprve do archivu (kde je převeden do balíčku AIP), a teprve poté je zpřístupňován uživatelům (převodem z balíčku AIP do balíčku DIP). Balíček DIP je podle této normy vždy odvozeninou balíčku AIP, nikoliv balíčku SIP. Balíček SIP (AIP, DIP) musí obsahovat informace o tom, které jeho části tvoří informační obsah (objekt CDO a technická metadata; např. jeden soubor v PDF a vybrané elementy PREMIS popisující jeho digitální identitu) a které jeho části tvoří archivační informace (tj. archivační metadata, např. v PREMIS).

Intelektuální entitu, která je reprezentována objektem CDO uloženým v balíčku SIP, označíme za digitální originál. V této práci se primárně zabýváme situací, kdy je objekt CDO v digitalizaci vytvářen tak, aby na jejím konci byl uložen do balíčku SIP již v archivačním formátu, a tedy archiv na něm nemusel vykonávat formátovou normalizaci (v průběhu zpracování balíčku SIP ve funkčním celku Příjem).

Intelektuální entita je v daný okamžik vždy reprezentována aktuálním digitálním objektem CDO, který je uložen v určitém formátu na určitém datovém nosiči, a to buď v balíčku SIP, nebo AIP, nebo DIP. Objekt CDO před jeho finalizací do podoby uložené v balíčku SIP je jiným případem (jde o nekompletní nebo nedokončený objekt CDO). Autenticita aktuálního objektu se posuzuje na základě vztahu aktuálního objektu CDO k jeho (stanovenému a popsánému) referenčnímu objektu.

### **6.3.2 Referenční objekt**

Za referenční objekt autenticity se považuje celek těchto dílčích objektů:

- fyzická předloha v určený okamžik času;
- digitální originál;
- popis vztahu mezi předlohou a digitálním originálem.

Bližší specifikace tohoto komplexního modelu referenčního objektu bude předmětem následující kapitoly. Součástí digitálního originálu musejí být klíčové vlastnosti. Tyto vlastnosti musejí být zachovány napříč všemi procesy odvozování a při každé reprodukci informačního obsahu (tj. jako vlastnosti intelektuální entity). Digitální originál se bude v průběhu času měnit, ale jeho klíčové vlastnosti musejí být zachovány, přičemž hlavním

rizikem jsou transformace (odvozování, jehož výsledkem jsou bitové změny objektu CDO). Při replikaci musí být zachována bitová totožnost (což implikuje zachování všech vlastností, nikoliv jen klíčových).

### 6.3.3 Složky autenticity

Z modelu InterPARES přejímáme rozdělení autenticity dokumentu na identitu a integritu dokumentu následovně. Identita intelektuální entity jsou atributy, které ji jednoznačně charakterizují a odlišují od ostatních. Pro její popis jsou potřeba údaje, které v modelu OAIS odpovídají identifikačním informacím a první části provenienčních informací (těm, které popisují původ a vznik digitálního originálu a jeho klíčové vlastnosti).

Integrita je vlastnost úplnosti a neporušenosti intelektuální entity, přičemž neporušenost neznamena zachování bitové totožnosti digitálního originálu, ale zachování jeho klíčových vlastností digitálního originálu, a to při odvozování i při reprodukci. Pro její popis jsou potřeba informace, které v modelu OAIS odpovídají informacím o neporušenosti a druhé části provenienčních informací (těm, které zahrnují popis historie změn uloženého informačního obsahu). Historie změn zahrnují dva typy informací – popis předcházejících objektů (generací) a popis událostí vykonávaných na těchto objektech. Tyto události dále rozdělujeme na odvozování (ty, které vytvářejí novou generaci objektů) a související události (tj. takové, které nemění objekt, např. kontrola neporušenosti prostřednictvím digitálního otisku).

Integritu intelektuální entity dále dělíme na integritu odvozeniny a integritu reprodukce. Integrita odvozeniny je otázkou historie objektu CDO a zahrnuje rizika ztráty klíčových vlastností zejména v procesu formátových konverzí. Integrita reprodukce je otázkou aktuálního zpřístupnění (formou prezentace v digitální knihovně) a zahrnuje především rizika ztráty klíčových vlastností při zobrazení, která představuje různá kvalita zpřístupňovacích softwarových aplikací.

Na úrovni objektu CDO můžeme rovněž hovořit o identitě a integritě. Digitální identita je klíčová pro technologickou interpretaci objektu CDO. Interpretační informace lze označit za popis digitální identity. Nejzřejmějším příkladem digitální identity je formát. Pro zjištění identity formátu je potřebná jednoznačná identifikace formátu, která dokáže odlišit různé verze nebo podtypy formátu. Na základě přesného zjištění digitální identity je možno vykonávat archivační opatření (např. nová verze formátu je vhodným archivačním formátem, zatímco stará verze nikoliv – to nelze odlišit na základě pouhého názvu formátu). Integrita formátu pak znamená, že formát odpovídá své specifikaci, tedy je autentický (je tím

formátem, za který se vydává). K ověření integrity (např. validace formátu) objektu je nejprve nutno znát jeho identitu – tento soubor se vydává za soubor ve formátu TIFF verze 6 (deklarovaná identita), je opravdu autentickým formátem TIFF verze 6? (Je vytvořen v souladu s oficiální specifikací formátu TIFF?). To ukazuje na životaschopnost pojmu identita a integrita nejen na úrovni zobrazeného dokumentu (intelektuální entity), ale také na úrovni objektu CDO, která je předpokladem reprodukce.

Objekt CDO se v současné praxi skládá z více dílčích složek (obrazová data, strukturální data, textová data v ALTO XML získaná procesem), které nazveme digitální komponenty. Odlišují se od sebe jinými mechanismy zacházení. Jednotlivé digitální komponenty objektu CDO nazveme obrazová komponenta, strukturální komponenta a textová komponenta. V souladu s tímto rozdělením je dále možno uvažovat i o identitě a integritě těchto komponent: objekt CDO je autentický, pokud jsou všechny jeho komponenty správně identifikované (otázka jejich digitální identity), úplné (obsahují všechny jednotlivé soubory) a neporušené (tj. validní v souladu s digitální identitou, např. formátovou specifikací).

#### **6.3.4 Archivační a interpretační informace**

Archivační informace musejí být podle normy ISO 14721 uchovávány spolu s uloženým objektem CDO a přidruženými technickými metadaty reprezentujícími interpretační informace, případně přidruženým zpřístupňovacím softwarem. Slouží jako nezbytná složka dokladů o autenticitě. Je zřejmé, že tyto archivační informace uložené s informačním obsahem budou v průběhu času nabývat, jak budou doplňovány o nové provenienční informace (např. v souvislosti s formátovými konverzemi).

Interpretační informace jsou klíčové pro aktuální objekt CDO a lze je označit za popis digitální identity. Znalost co nejvíce informací o digitální identitě (formátu, barevném profilu apod.) je nezbytná pro možnost správné reprodukce i pro zjišťování rizik zastarávání.

Archiv může interpretační informace pro aktuální objekt CDO měnit, i když zůstává objekt CDO zachován, a to s ohledem na změny počítačového prostředí. Archiv například může vyměnit zpřístupňovací software, pokud se ukáže, že zastarává (např. není kompatibilní s nejnovějšími verzemi internetových prohlížečů), nebo doplnit údaje obsažené v technických metadatech (např. pokud pro formát objektu CDO předtím neexistoval identifikátor PUID a nyní jej registr PRONOM zavedl). Pro minulé generace objektů CDO, které se nezachovávají (např. mažou v průběhu digitalizace), přestávají být technická metadata důležitá z hlediska možnosti reprodukovatelnosti a stávají se součástí archivačních metadat (jsou důležitá jako součást provenience zachycující historii odvozování objektu CDO).

## 7 Specifikace požadavků autenticity pro digitalizáty knih

### 7.1 Stanovení referenčního objektu

Referenční objekt authenticity digitalizátů knih je celek těchto dílčích objektů: a) fyzická předloha v určený okamžik času, b) digitální originál (včetně popisu jeho klíčových vlastností), c) popis vztahu mezi fyzickou předlohou a digitálním originálem (úroveň abstrakce podle FRBR a popis odvozování). V této podkapitole se zabýváme typovým stanovením referenčního objektu. To znamená takovým stanovením, které se uplatňuje na všechny digitalizáty knih. Základní typ z hlediska intelektuální entity reprezentované digitálním originálem je digitalizát knihy.

Fyzickou předlohou se rozumí tištěná kniha (případem, kdy se digitalizují mikrofilmy tištěných knih, se zde nezabýváme) a okamžikem času buď její aktuální stav v době digitalizace, nebo předpokládaný stav knihy v době jejího vzniku.

Digitalizační projekt musí jako svůj cíl stanovit podobu digitálního originálu (tj. objektu CDO, který bude uložen do balíčku SIP jako výstup digitalizace a který reprezentuje intelektuální entitu), to znamená jeho typovou digitální identitu a jeho klíčové vlastnosti. V kontextu řízení celého životního cyklu jednou organizací je digitální originál stanoven současně jako cíl digitalizačního projektu i jako předmět následného dlouhodobého uchovávání v archivu. Stanovení podoby digitálního originálu zohledňuje nejen kvalitu produkce, ale také potřeby dlouhodobého uchovávání. Například formát DNG splňuje požadavky kvality, ale nikoliv požadavky na archivační formát.

Typová digitální identita se vztahuje na technické parametry digitálního originálu, což znamená specifikaci výsledného objektu CDO. Typová digitální identita zahrnuje zejména formáty jednotlivých digitálních komponent (např. formát TIFF pro obrazovou komponentu) a požadované obrazové vlastnosti obrazových komponent (např. model RGB, bitová hloubka 24 bitů). Klíčové vlastnosti digitálního originálu zahrnují jednak převáděné vlastnosti tištěné předlohy (to, co bude digitalizováno), jednak přidané digitální vlastnosti (to, čím bude digitální originál obohacen, např. plnotextová prohledatelnost). Rozdíl mezi typovou digitální identitou a klíčovými vlastnostmi je zřejmý – nelze zaručit, že typová digitální identita bude zachována v dlouhodobém horizontu (např. formát TIFF bude převeden v budoucnosti do jiného), zatímco klíčové vlastnosti by měly být zachovány napříč časem (např. plná barevnost).

Určení podoby digitálního originálu závisí také na určení úrovně abstrakce, která má být digitalizací zachycena (z hlediska modelu FRBR je knihu možno pojímat jako vyjádření, manifestaci nebo exemplář). Podle této úrovně se musí stanovit způsob odvozování, jehož výsledkem bude digitální originál požadovaných vlastností.

### 7.1.1 Úroveň vyjádření

Na úrovni vyjádření jde v případě knih o textové vyjádření. Klíčovými vlastnostmi bude čistě textový obsah ve své struktuře. Ilustrace v knize, které jsou dílem jiného autora, nejsou součástí textového vyjádření, a tedy je není potřeba převádět.<sup>77</sup> Podoba digitálního originálu pak může být stanovena například jako soubor v strukturovaném textovém formátu (např. TEI) nebo obrazové soubory ve formátu TIFF pro každou stránku či jeden soubor formátu PDF vytvořený z textového souboru. Časový aspekt fyzického originálu bude tvořit kniha v době svého vzniku (tj. bude se odhlížet od textových prvků přidaných při správě knihy, např. názvu knihovny v otisku razítka na stránce knihy).

Dále musí být stanoven způsob odvozování. Je možné, že kniha bude zdigitalizována tak, že ji specialista přepíše přímo do formátu TEI. Nebo je možné, že bude snímána jako obrazová data, na kterých bude následně provedeno OCR a výsledný text specialista zkontroluje a přepíše do formátu TEI. Způsoby odvozování tedy mohou být různé, ale musejí být stanoveny.

Při posuzování autenticity digitalizátů převáděných na úrovni textového vyjádření pak může hrát roli zvážení toho, která ze způsobů bude posuzovatel autenticity považovat za nejspolehlivější z hlediska věrného převodu textového obsahu knihy. K tomuto posuzování může napomoci i znalost zajištění praktické realizace projektu (např. poměr počtu pracovníků k počtu digitalizovaných knih nebo úroveň jejich zaškolení v kódování textu do formátu TEI), pokud je ovšem pracovní postup projektu řádně zdokumentován. Jakoukoliv chybu v textu lze považovat za narušení autenticity tohoto typu digitalizátu. Text musí být převeden v úplnosti a své struktuře.

V současné praxi není postup digitalizace knih na úrovni vyjádření užíván (jde však o postup někdy uplatňovaný při digitalizaci rukopisů). Popis tohoto možného přístupu nám posloužil jako názorný příklad toho, že referenční objekt musí zahrnovat nejen otázku fyzické předlohy a jejího časového určení a stanovení podoby digitálního originálu, ale také popis

---

<sup>77</sup> Specifickým případem jsou ilustrace nebo jiné netextové prvky, které jsou součástí vyjádření, např. vlastní ilustrace autora textu. Tento a podobné specifické případy opomíjíme.

zvolené úrovně abstrakce digitálního převodu a postupu odvozování. Rovněž slouží jako ukázka toho, že ve fázi digitalizace knih je otázka autenticity (ve smyslu věrné digitální reprezentace) v první řadě otázkou stanovení referenčního objektu. Textový digitalizát v souboru formátu TEI může být zcela autentickou reprezentací knihy jako textového díla a jako digitální intelektuální entita pro většinu typů užití rovnocenný s tištěnou knihou (jakýkoliv text je ostatně formou nebinárních digitálních informací, které lze podobně jako bity dokonale přenášet a kopírovat).

### 7.1.2 Úroveň manifestace (vydání) a exempláře

V současné praxi masové digitalizace knih je zachycovanou úrovní abstrakce tištěné předlohy úroveň manifestace (vydání) i exempláře (zachycením exempláře knihy se současně zachycuje úroveň vydání). Tato dvojí snímaná úroveň má analogii v popisu tištěných knih v katalogizační praxi. Přesto lze obvykle z kontextu řízení životního cyklu digitalizátů v dané organizaci odvodit, že upřednostňovanou úrovní je úroveň vydání. Například pravidla pro poskytování dotací z podprogramu VISK7 vyžadují, aby nebyl digitalizován exemplář vydání, které již bylo digitalizováno: argument, že daná knihovna má jiný exemplář (téhož vydání), než je ten, který byl již zdigitalizován v nějaké předchozí české digitalizaci, nelze uplatnit.<sup>78</sup> Z hlediska zpřístupňování v digitálních knihovnách platí, že knihy v nich lze zpravidla vyhledávat podle bibliografických údajů vztahujících se k vydání předlohy, nikoliv k exempláři.<sup>79</sup> Mohou se však vyskytovat projekty, které budou akcentovat specifčnost exemplářů (např. poznámky vepsané do knihy jejich vlastníky, kteří jsou považováni za významné osobnosti). Pro zdůraznění této úrovně exempláře je však potřeba, aby knihovna zavedla specifický bibliografický popis, který bude zachycovat tyto prvky exemplářů.

Časovým určením fyzické předlohy je v současné praxi stav tištěné knihy v době její digitalizace. Pokud by za časové určení byla zvolena kniha v době svého vzniku, pak by fyzická předloha byla konceptuálním objektem (vytvořeným na základě výzkumu tehdejší podoby knihy). Takovýto přístup by vyžadoval odborné digitální restaurování barev a odstranění přidaných prvků (razítek apod.). Tento přístup je nezvyklý v kontextu digitalizace knih, ale běžný pro digitalizaci filmů ve filmových archivech, například v současném projektu Národního filmového archivu (92). Spor o způsob takového

---

<sup>78</sup> Viz „seznam všech dokumentů navržených k reformátování... musí být ověřen z hlediska duplicit v Registru digitalizace.“ (48, s. 6) Tento registr pomáhá identifikovat duplicitu na základě identifikátoru vydání knihy.

<sup>79</sup> Viz např. digitální knihovna Kramerius (<http://kramerius4.nkp.cz/search/>), kde lze vyhledávat podle těchto údajů: ISBN/ISSN; název titulu; klíčová slova; autor, rok vydání, MDT, DDT.



restaurování, který tento archiv vede s Asociací českých kameramanů, ukazuje na odlišné přístupy k stanovení autentické digitální reprezentace filmu, v jejichž jádru je právě spor o konceptuální objekt (o to, jak film vypadal v době svého vzniku).

Typová digitální identita digitálního originálu musí být specifikována jako digitální identita tří digitálních komponent objektu CDO: obrazové komponenty (rastrový archivační formát a základní obrazové vlastnosti, např. prostorové rozlišení a bitová hloubka), textové komponenty (formát ALTO XML) a strukturální komponenty (formát METS XML). Proto budeme tento způsob digitalizace knih, který je nejběžnějším způsobem současné praxe, pro zjednodušení nazývat rastrová digitalizace.

Pro referenční objekt rastrové digitalizace je dále zapotřebí stanovit způsob odvozování z předlohy do digitálního originálu. Tento způsob odvozování, spolu s určením času fyzické předlohy a úrovně abstrakce (exemplář / vydání), je základem pro posouzení míry věrnosti digitalizace, a tedy autenticity digitální reprezentace ve vztahu k její předloze. Například skutečnost, že se provádějí ořezy uvnitř hrany (tj. část informačního obsahu tištěné knihy se digitálně nepřevádí), může být důvodem ke konstatování nižší míry věrnosti, a tím i nižší míry autenticity.

### **7.1.3 Stanovení klíčových vlastností**

Klíčové vlastnosti digitálního originálu zahrnují jednak převáděné vlastnosti tištěné předlohy (to, co bude digitalizováno; např. obrazové a strukturální vlastnosti), jednak přidané digitální vlastnosti (to, čím bude digitální originál obohacen). Tyto vlastnosti by měly být stanoveny před zahájením digitalizačního projektu jako cíl digitalizace a měly by být součástí popisu digitalizačního projektu jako deklarace klíčových vlastností, které budou dlouhodobě uchovávány. V současné praxi rastrové digitalizace se však s jejich stanovením nesetkáváme, a to ani ve fázi digitalizace, ani v následné fázi archivace, což představuje zásadní problém.

Klíčové vlastnosti musejí být definovány především na úrovni intelektuální entity (tedy jako vlastnosti toho, co může čtenář vidět v digitální knihovně při prezentaci). Mezi digitalizované klíčové vlastnosti může patřit: plná barevnost, dobrá čitelnost textu, původní posloupnost stran, úplnost stran a listovatelnost. Mezi přidané digitální vlastnosti například plnotextová prohledatelnost a kopírovatelnost textu. Definice klíčových vlastností intelektuální entity stanoví cíle, kterých je nutno dosáhnout v digitalizaci i při dlouhodobém uchovávání a zpřístupňování, a jejich popis by měl být srozumitelný cílové komunitě všech čtenářů – jako deklarace toho, co bude dlouhodobě uchováváno a prezentováno v digitální knihovně, a tedy co mohou čtenáři od digitalizátů knih vyžadovat. Na základě definice

klíčových vlastností intelektuální entity pak lze stanovit takovou typovou digitální identitu objektu CDO, která umožňuje zaznamenat tyto klíčové vlastnosti, a to současně s přihlédnutím k potřebám digitální archivace. Plná barevnost například vyžaduje užít minimální bitovou hloubku 24 bitů.

Obecná definice klíčových vlastností pro celý digitalizační projekt, resp. pro různé digitalizace téže knihovny, může posloužit jako náhrada zápisu těchto vlastností do metadat, protože tyto vlastnosti nelze automatizovaně kontrolovat. Tyto obecné definice klíčových vlastností však lze (a je nutné) zapsat do dokumentace digitalizačního projektu nebo programu dané organizace a také do dokumentace archivu popisující pravidla pro uchovávání (jako deklarace toho, co musí být zachováno).

Klíčové vlastnosti intelektuální entity je dále možno a záhodno specifikovat konkrétnějšími technickými parametry (tedy na úrovni objektu CDO, která není vnímatelná), které lze automatizovaně ověřit a které mají být zachovány napříč transformacemi. Příkladem takové vlastnosti je matematicky bezeztrátová obrazová komprese. Tu je vhodné zapsat do výše uvedené dokumentace i do metadat obrazových dat a při následné konverzi do jiného formátu na základě tohoto údaje v metadatech provést automatizovanou kontrolu (získáním informací z nových obrazových souborů extraktorem metadat a srovnáním s požadovanou specifikací).

Vzhledem k tomu, že neexistuje široce užívané metadatové rozšíření standardu PREMIS pro zápis signifikantních vlastností pro digitalizáty knih (resp. obecně pro rastrová data), má však zásadní význam to, co bylo uvedeno výše: zapsat klíčové vlastnosti v obecně srozumitelné podobě do dokumentace organizace. Takovéto obecnější vymezení pak umožní v průběhu času, až bude dostupné rozšířené schéma pro zápis signifikantních vlastností pro digitalizáty knih, nové elementy využít a doplnit do nich tyto údaje.

## **7.2 Identita digitalizátu knihy jako intelektuální entity**

Pro určení identity digitalizátu knihy jsou klíčové identifikační a některé provenienční informace. Autenticita digitalizované knihy jako intelektuální entity především znamená, že digitalizát knihy je tím, za co jej vydává organizace ve své digitální knihovně, to znamená, jaké identifikační a provenienční informace prezentuje spolu s knihou.

Atributy identity digitalizátu knihy musejí vycházet z referenčního objektu. Otázkou je, jaké všechny atributy jsou potřebné pro vyjádření identity a musejí být zapsány do přidružených metadat nebo dokumentace a uchovávány spolu s digitalizátem jako součást

požadavků autenticity. Z hlediska atributů tištěné předlohy je vždy důležitá identita exempláře a identita vydání tištěné knihy. V tomto ohledu je samozřejmě základním prostředkem vyjádření identity katalogizační záznam, který popisuje úroveň vydání i exempláře. Ten však v současné praxi představuje řadu problémů, o kterých pojednávají následující dva oddíly.

### **7.2.1 Identita exempláře knihy jako zdroje digitalizátu**

Nejprve se budeme věnovat problémům vyjádření identity exempláře. Zaprvé jde o identifikátor exempláře. Signatury se mohou měnit a původní nemusejí být zachovány (např. přelepením signatury). Jedinečné čárové kódy vytvořené pro digitalizační projekt teoreticky mohou být vhodným identifikátorem, pokud ovšem nejsou po skončení digitalizace skartovány (případ, kdy se vkládají do knihy na samostatném listu). Pokud se vlepují do knihy a zachovávají se, pak v případě dalšího digitalizačního projektu musí být vytvořena nová sada čárových kódů, která musí zachovat jedinečnost ve vztahu k předchozím užitým sadám. To může být organizačně náročné. Jistým řešením může být vytištění pořadového čísla exempláře do knihy, které ve spojení s perzistentním identifikátorem vydání umožní jednoznačné určení exempláře (např. vydání knihy v pěti exemplářích bude mít čísla 1 až 5). V případě vyřazení některého exempláře (např. při destruktivní digitalizaci) ale musí být zachováno pořadí pro případné další akvizice (tj. aby např. další pořízený exemplář dostal pořadové číslo „6“). Optimálním řešením se může zdát zavedení akvizice takovým způsobem, aby se do knihy dalo razítko s datem zpracování a v případě, že tentýž den bylo zpracováno více knih, přidá se k němu pořadové číslo platné pro daný den. Tímto způsobem by z identifikátoru exempláře bylo možno zjistit i datum akvizice. Předpokládá však změnu akviziční praxe a může být problém opatřovat takovým razítkem knihy již ve fondech uložené. Otázkou také je, nakolik je vlastně takováto jedinečná identifikace exempláře potřebná. K digitalizátu knihy, jehož předloha bude později z fondu vyřazena, již stejně nebude možno exemplář dohledat, jakkoliv dobře by byl identifikován. V tomto ohledu je tedy možné spokojit se se zaznamenáváním identifikátoru exempláře, který byl užíván v době digitalizace (např. signaturou), do metadat.

Druhým nezbytným atributem exempláře je název vlastníka (zpravidla tedy knihovny, která exemplář vlastní). Tento údaj je důležitý nejen z hlediska autenticity (provenience), ale také z právních důvodů (např. majetková práva). Třetím atributem je informace o stavu předlohy v době digitalizace (např. stav: nekvalitní předloha). Ten je důležitý pro rastrovou digitalizaci zachycující aktuální stav exempláře, a sice jako komentář ke kvalitě obrazu. V praxi bývá do metadat zaznamenáván.

Dalšími atributy exempláře knihy pak mohou být informace o tom, jakým způsobem byla kniha získána do fondu, případně o tom, že bude po digitalizaci vyřazena (případ destruktivní digitalizace). První údaj je důležitý jako informace pro čtenáře, která mu pomáhá porozumět tomu, proč kniha zobrazená v digitální knihovně obsahuje razítko jiné knihovny (které je v ní z toho důvodu, že kniha byla získána z jiné knihovny, která ji vyřadila), což není výjimečný jev.<sup>80</sup> Druhý údaj je důležitý jako informace o tom, že exemplář knihy, jejíž digitalizát je vystaven v digitální knihovně, již nelze dohledat, a tedy není možné srovnat jej s tímto digitalizátem.

Specifickým případem je kompletace digitalizace z více exemplářů, například pokud se při kontrole kvality zjistí, že předloha nebyla úplná, a chybějící strany se doplní digitalizací jiného exempláře. Tento přístup se sice z praktického hlediska zdá být přirozeným, ale z dlouhodobého hlediska přináší více problémů než užitku. Údaj o více užitých exemplářích se v praxi do metadat nezaznamenává, nehledě na obtížnost takového zápisu. Absenci záznamu této skutečnosti lze považovat za nedostatečný popis identity. Odlišné části se také mohou významně lišit z hlediska kvality (jiný stupeň opotřebení apod.). Obojí samozřejmě snižuje míru autenticity digitalizátu. Nelze proto než doporučit raději provést opětovnou digitalizaci celého kompletního zdroje.

### **7.2.2 Identita vydání knihy jako zdroje digitalizátu**

Pro vyjádření identity vydání slouží v běžné praxi bibliografický záznam. Klíčovým prvkem v tomto ohledu je perzistentní identifikátor vydání. Takovým je identifikátor ISBN, který ale není přidělován starším knihám (vydaným před zavedením národního systému ISBN v dané zemi). Systém perzistentní identifikace vydání musí být zaveden před zahájením digitalizačního projektu. Tomuto problému se blíže věnuje Kapitola 9.

Z hlediska bibliografického záznamu je také důležité, aby byl co možná nejobsáhlejší. Je třeba mít na paměti, že digitalizát knihy může jednou zcela nahradit tištěné předlohy, které se vlivem degradace rozpadnou. Druhou podmínkou je, aby byl bibliografický záznam kvalitní. Před digitalizací konkrétní knihy musí být pečlivě zkontrolován a doplněn, a to s exemplářem v ruce. Optimálním řešením je, aby si digitalizační projekt zajistil dostatek zdrojů na rekatalogizaci knih. Špatná identifikace předlohy digitalizátu znamená narušení možnosti ověřit jeho identitu z hlediska předlohy, a tím narušení autenticity. Čtenář bude

---

<sup>80</sup> Jde o běžný případ digitalizace projektu NDK.

uváděn v omyl, když mu bude prezentován digitalizát spolu se zavádějícím nebo chybným popisem.

Dalším klíčovým aspektem identity vydání knihy je, aby byly zaznamenány veškeré údaje o barevnosti předlohy, zejména v případě, že součástí digitalizačního projektu je digitalizace části fondu ve stupních šedi. Takovéto informace jsou důležité pro posouzení autenticity z hlediska barevné věrnosti: pokud projekt deklaruje věrnou barevnost, čtenář z prezentované digitalizované knihy samotné, bez přidružených identifikačních informací, většinou nemůže poznat, zda kniha opravdu neobsahovala žádné barevné prvky (a tedy digitalizace ve stupních šedi nenarušuje věrnost digitalizačního převodu). Problém je, že záleží na dané katalogizační politice, do jaké míry a jakým způsobem standardizuje záznam barevných vlastností. Současná pravidla NK ČR například neposkytují žádné pokyny k záznamu barevných informací o textu tištěné knihy (např. užití černého a červeného textu pro odlišení jiných částí knihy), zatímco pro elektronické publikace je taková možnost uvedena (93).

### **7.2.3 Identita digitalizátu knih jako intelektuální entity**

Prvním nezbytným atributem identity digitalizátu knihy je skutečnost, že jde o digitalizovanou knihu (tj. digitální reprezentaci tištěného dokumentu). Čtenář nesmí být uveden v omyl, že je mu prezentována digitálně nativní kniha (e-born publikace).

Další atributy pak musejí odlišovat digitalizát knihy od jiných možných digitalizátů téhož zdroje, ať již existujících nebo předpokládaných, a to jak z hlediska zpřístupňování uživatelům, tak správy v archivu. Různé digitalizace se budou lišit, což vyjadřuje koncept referenčního objektu, který byl navržen v této dizertační práci. Atributem identity digitalizátu knihy je vlastník digitalizátu (tedy knihovna, která vlastní tištěnou předlohu), digitalizační projekt (jedna knihovna může mít v průběhu času více digitalizačních projektů a stejná kniha může být zdigitalizována stejnou organizací vícekrát v průběhu času) a datum vzniku digitalizátu (teoreticky může v rámci jednoho projektu vzniknout více digitalizátů stejné knihy – odlišné snímání, byť v rámci téhož nastavení, zakládá nový digitalizát). Identitou digitalizátu je také (v danou dobu) její URL lokace, která se však mění.

## **7.3 Identita a integrita objektu CDO**

Integritu intelektuální entity je třeba pojímat jako identitu a integritu objektu CDO. Identitu objektu CDO (tj. digitální identitu) tvoří v současné praxi obrazová, textová

a strukturální komponenta a jejich struktura. Každá tato digitální komponenta je tvořena skupinou jednoho nebo více souborů, které se liší formátem, případně dalšími vlastnostmi. Součástí identity všech souborů je jejich název a relativní umístění v balíčku, které tvoří základní strukturu objektu CDO. Rozdíl mezi strukturou objektu CDO a strukturální komponentou objektu CDO spočívá v tom, že strukturální komponenta zaznamenává informace o struktuře, zatímco strukturu objektu CDO tvoří jednotlivé komponenty a názvy jejich souborů a umístění, které jsou popsány ve strukturální komponentě. Integritu struktury objektu CDO tak tvoří úplnost těchto souborů a jejich správné pojmenování a umístění podle stanovené podoby digitálního originálu. Hlavním mechanismem pro ověření neporušenosti jednotlivých souborů je digitální otisk. Ten musí být vytvořen bezprostředně po vytvoření souboru (nebo jeho přejmenování) a zaznamenán do metadat společně s informacemi o datu vytvoření otisku a užitém algoritmu. Pokud by byl digitální otisk vytvořen později, pak již nelze zaručit, že v této pozdější době nedošlo k porušení souboru (pak by byl otiskem porušeného souboru, a tedy nemohl plnit svoji funkci). Integrita aktuálního objektu CDO musí být tedy ověřena tím, že jsou všechny jeho soubory přítomné (úplnost), uložené a pojmenované podle daného datového modelu (neporušenost struktury objektu CDO) a nezměněné (neporušenost souborů), a dále tím, že je zachována specifická integrita jednotlivých komponent, tedy je v souladu s jejich stanovenou identitou.

Typovou digitální identitu všech komponent objektu CDO tvoří specifický formát a v případě obrazové komponenty základní obrazové vlastnosti. Formátem obrazové komponenty je v současné praxi nejčastěji TIFF (verze 6) a JP2; formátem textové komponenty ALTO XML a formátem strukturální komponenty METS XML. Obrazovou i textovou komponentu zpravidla tvoří skupina souborů (jeden soubor pro jednu stránku), strukturální komponenta je uložena v jediném souboru. V tomto souboru jsou však zaznamenány nejen strukturální informace o umístění digitálních komponent, ale též interpretační a archivační informace vnořené do jiných metadatových schémat v rámci schématu METS a odkazy na jiné metadatové soubory obsahující technické a archivační informace. Tato skutečnost tedy musí být explicitně a dostatečně přesně popsána v dokumentaci užitého aplikačního metadatového profilu.

Identitu formátu tvoří zejména jeho jednoznačná identifikace a formátový profil. Pro jednoznačnou formátovou identifikaci lze užít pouze identifikátor PUID, což předpokládá užití nástroje pro formátovou identifikaci, který dokáže pracovat s údaji registru PRONOM. Výstup formátové identifikace se zapisuje do technických metadat jako součást interpretačních informací. Vhodnou praxí je doplnkově uvádět též oficiální název formátu

a jeho verzi, k zjištění těchto informací ze souborů slouží extraktory metadat (nejužívanější je JHOVE).

Formátový profil znamená volbu parametrů v rámci daného formátu, přinejmenším nastavení komprese, v případě formátu JP2 pak několik specifických hodnot (některé jsou důležité z hlediska způsobu užití tohoto formátu v prezentačním systému). Pro zjištění informací o profilu je nutno užít extraktory metadat a tyto informace zaznamenat do technických metadat. Pro identifikace formátů textové a strukturální komponenty je nutné určit nejen verzi formátu XML, ale také verzi schématu METS a formátu ALTO.

Identitu obrazové komponenty dále představují atributy obrazové specifikace, které nejsou svázány s konkrétním formátem. Mezi hlavní patří bitová hloubka, barevný model, informace o typu komprese (matematicky bezeztrátová, vizuálně bezeztrátová apod.) a barevný profil.

Posouzení integrity digitálních komponent objektu CDO znamená posouzení toho, zda jsou v souladu s jejich identitou. Integrita formátu (označovaná jako formátová validace) znamená, že formát je ve shodě se svou oficiální specifikací. Toto posouzení však závisí na konkrétním nástroji, jeho verzi a jím užitě definici validace. Integrita profilu nebo obrazových vlastností v první řadě znamená, že komponenta je v souladu s tím, jak byla stanovena při typové definici digitálního originálu.

Identita objektu CDO se bude v průběhu času nutně měnit v souvislosti se změnou technologického prostředí a potřebou provádět v důsledku této změny formátové konverze. V mezidobí však musí být zachována integrita objektu CDO, a to jak na úrovni základní struktury (názvy a umístění souborů), neporušenosti a úplnosti souborů, tak na úrovni identity jednotlivých komponent. Digitální identita musí být vždy podrobně specifikována. Podle této specifikace pak musí být soubor s danými atributy identity (formát, barevný model apod.). Při digitalizaci to znamená stanovit typovou identitu vytvářeného digitalizátu. Určit za formát obrazové komponenty TIFF, aniž by byla určena také jeho verze, nebo formát JP2, bez toho, aby byl určen jeho konkrétní profil, tedy není dostatečné.

## **7.4 Řetězec odvozování objektu CDO**

### **7.4.1 Tištěná předloha jako první objekt CDO**

Tištěnou knihu, která je předmětem digitalizace, je nutno v řetězci odvozování považovat za první objekt CDO (fyzický objekt CDO). Je součástí referenčního objektu při

jakémkoliv každém typu digitalizace. V současné praxi rastrové digitalizace lze určit jako rozhodné otázky přípravy tištěných knih pro digitalizaci následující.

V první řadě musí být vybrán takový exemplář vydání, který je kompletní. Kniha tedy musí být předem pečlivě zkontrolována. Předloha pak musí být vyčištěna (např. od prachu) a upravena tak, aby proces snímání probíhal hladce (např. vyrovnaním zahnutých stran). Nedostatky předlohy (např. výrazný stupeň degradace nebo poničení) musejí být zaznamenány do metadat. Pokud projektová dokumentace nestanovila metodu digitálního restaurování, nelze v procesu vytváření digitálního originálu provádět úpravy za účelem dosažení předpokládané podoby v době vzniku knihy (např. digitálně odstraňovat kaňky), neboť by šlo o rozpor se stanoveným postupem, a tedy narušení autenticity.

## **7.4.2 Vytváření digitálního originálu**

### **7.4.2.1 Původní snímky**

Rastrová digitalizace je vždy nutně ztrátový převod – digitálně lze převést pouze obrazové a strukturální vlastnosti fyzické předlohy. Textová digitalizace (zachycování předlohy na úrovni vyjádření) může být zcela bezztrátový převod, pokud je převeden veškerý text a jeho struktura (např. odstavce).

Kvalita původních snímků je nutně determinovaná možnostmi aktuální snímací technologie. Svoji roli také hrají ekonomické faktory (cena snímacích zařízení a úložných nosičů). Dále je třeba mít na paměti, že některé volby snímání nevratně ovlivňují podobu digitálního originálu<sup>81</sup> a že vždy záleží na typu konkrétního zařízení, jeho kvalitě a možnostech nastavení (včetně možnosti zrušit případné výchozí nastavení, pokud nevyhovuje).

Rozhodným prvkem digitalizačního projektu je volba takových snímacích zařízení, která dokáží vytvářet obrazová data v požadovaných parametrech. Základním požadavkem současné praxe je schopnost ukládat původní snímky v nekomprimovaném rastrovém formátu (RAW nebo TIFF). Jiná volba není v současnosti vhodná. Dalším potřebným parametrem je schopnost zařízení ukládat do původních snímků metadata ve formátu EXIF. Tato metadata jsou důležitá pro autenticitu (viz oddíl 7.8).

---

<sup>81</sup> Z původních snímků ve ztrátové kompresi nebo s nízkým rozlišením již nelze v následném zpracování vytvořit digitalizát ve vyšší kvalitě.



Na tomto místě je také zapotřebí upozornit na to, že paušálně stanovené prostorové rozlišení pro všechny knihy projektu znamená, že reálné rozlišení konkrétních digitalizátů (a tedy míra čitelnosti písma) se bude knihu od knihy přinejmenším nepatrně lišit (v závislosti na velikosti písma v předloze). Při fotografování se prostorové rozlišení nenastavuje (jde čistě o otázku kvality fotoaparátu a objektivu a možnosti přiblížení a zaostření). Při tomto typu snímání je však nutno prostorové rozlišení vypočítat (na základě údajů o velikosti knihy a pixelové velikosti původního snímku) a zaznamenat do obrazového souboru (to je užitečné zejména pro možnost budoucího vytištění v původní velikosti).

#### **7.4.2.2 Transformace do digitálního originálu**

Digitalizační projekt musí stanovit všechny kroky odvozování do výsledného digitálního originálu v určené podobě, a to pro všechny tři digitální komponenty.

Odvozování struktury objektu CDO a strukturální komponenty se obvykle odvíjí podle posloupnosti snímání a odpovídajícího pojmenovávání souborů. Systém pojmenování musí být vymyšlen tak, aby umožnil vytvoření požadované struktury objektu CDO, případně umožnil odlišit všechny generace obrazových komponent. Zachování všech těchto generací do doby, než bude provedena konečná kontrola digitálního originálu, může být vhodným způsobem pro odhalování příčin případných chyb.

Při odvozování obrazové komponenty je nutno určit rozhodný počet generací, který bude i předmětem metadatového popisu. Některé transformace podobného typu lze (nebo je nutné vzhledem k užití technologii) vykonat najednou, takže může být (nebo musí být) stanoveno, že za novou generaci se považuje až výsledek všech těchto transformací. Například při zpracování formátu Canon RAW v aplikaci výrobce tohoto fotoaparátu lze jednotlivé transformace (po jejich nastavení na obraze) okamžitě uložit do souboru (např. nejprve se provede natočení snímku a uloží se, poté se provede jeho ořez a uloží se). Tím vznikají dvě nové generace obrazové komponenty. Aplikace Adobe pracují s obrazovými soubory ve formátu DNG tak, že transformace se pouze nastavují (toto nastavení se ukládá do konfiguračních souborů aplikace, nikoliv do souboru) a do souboru se ukládají až současně s konverzí (exportem) do formátu TIFF. Je otázkou smysluplnosti, kolik generací se v prvním případě zvolí (zda natočení i ořez považovat za jednu transformaci, nebo dvě). Tato volba však musí být zohledněna v metadatech (bude se popisovat buď jeden nový, nebo dva nové objekty). Podobně je tomu v případě, kdy se do formátu JP2 zapisují metadata do dvou odlišných částí (boxů) tohoto formátu (to je možné je považovat za jednu transformaci).

Transformace obrazových komponent lze zařadit do dvou skupin: formátové konverze a intraformátové transformace. Formátové konverze zahrnují i konverzi do meziformátu. V současné praxi můžeme rozeznat tyto možné kombinace řetězce formátových konverzí: TIFF (snímkový formát a výsledný archivační formát jsou jeden a tentýž formát); TIFF – JP2; RAW – TIFF; RAW – DNG – TIFF; RAW – TIFF – JP2; RAW – DNG – TIFF – JP2.

Z hlediska intraformátových transformací můžeme v současné praxi rozlišit tyto typy a podtypy (seznam není vyčerpávající): rekonstrukce obrazu (ořez, vyrovnaní zkosení, vyrovnaní zakřivení), obrazové úpravy (barevné korekce, úpravy tonality, doostření), převod do jiného barevného profilu; vložení metadat do formátu JP2 (vložení metadat do části „XML Box“, vložení metadat do části „UUID Box“). Uvedené typy transformací zahrnující podtypy lze pojímat souhrnně jako jednu transformaci, tj. stanovit, že nová generace vzniká pouze po dokončení všech dílčích transformací. Tomu pak musí odpovídat metadatový zápis v PREMIS. V tomto případě bude výslednému objektu připsáno několik událostí, u kterých by mělo být také zaznamenáno pořadí, ve kterém se uskutečnily. V některých případech (viz uvedený příklad zpracování obrazu v aplikacích Adobe) intraformátové transformace vytváří novou generaci až při exportu z dané explikace (tj. současně s formátovou konverzí).

Specifickou větví současné praxe, která se nevyskytuje ve všech projektech, je odvozování prezentačních variant obrazových komponent při produkci. Prezentační varianta může být odvozována ze stejné předcházející generace jako obrazová komponenta v archivačním formátu (např. z formátu TIFF se po provedení případných intraformátových transformací vytváří obrazové komponenty v archivačním formátu JP2 v matematicky bezztrátové kompresi i prezentačním formátu JP2 ve ztrátové kompresi), nebo z obrazových komponent v archivačním formátu.

Při odvozování textové komponenty může být jako zdroj užít dočasný extrakční meziformát, který je vytvořen pouze za účelem provedení procesu OCR, a poté je smazán. Pro zachování funkce prohledatelnosti je nutné, aby při odvozování nových generací obrazových komponent byla zachována pixelová velikost obrázku; v opačném případě nebude mapování ve formátu ALTO XML na daný obrazový soubor odpovídat. Proces OCR se tak musí provádět až po takových transformacích, které mění pixelovou velikost souboru (např. ořez).

Strukturální komponenta (část souboru METS XML popisující strukturální mapu) a textová komponenta mají pouze jednu generaci a jedinou událost (vytvoření).

Pro všechna odvozování obrazové a textové komponenty konkrétního objektu CDO musí být zachován princip, že prvním objektem na počátku řetězce odvozování je vždy tentýž původní snímek. Není možné vytvářet prezentační variantu nebo textovou komponentu odvozováním z jiného původního snímku (tj. z nového snímání téže knihy). Převod do vybraného barevného profilu musí být uskutečněn buď před snímáním (tj. v rámci nastavení skeneru), nebo při transformaci původního snímku.

Postup odvozování každé komponenty musí být předem zaznamenán a dodržován. Po každé transformaci pak musí být vytvořen digitální otisk a před zahájením následující transformace zkontrolován. Po každé transformaci musí být také zkontrolován například formát (tj. zda je validní).

#### **7.4.3 Dodávání digitálního originálu do archivu**

Akt dodávání dat do archivu je také odvozováním, které musí být pouze replikací. To znamená, že se mění jen nosič: z původního nosiče balíčku SIP (u producenta) na nosič dočasného úložiště funkčního celku Příjem (případně mezi těmito úseky na nosič, na kterém jsou balíčky SIP fyzicky doručeny do archivu). Základem ověření integrity je zde kontrola neporušenosti na základě digitálního otisku. Po vytvoření digitálního otisku pro všechny soubory digitálního originálu by měl být otisk ještě zkontrolován, než budou balíčky SIP dodány do archivu (jako kontrola správného vytvoření otisku). Specifickým požadavkem je užití digitálního podpisu jako mechanismu ověření původce dat. Pro řízení životního cyklu jednou organizací je však mechanismus zřejmě nepodstatný.

#### **7.4.4 Odvozování při archivaci**

V této práci se zabýváme případem, kdy digitální originál v balíčku SIP je již v archivačním formátu. Z tohoto důvodu při vytváření balíčku AIP ve funkčním celku Příjmu archivu OAIS není prováděna transformace objektu CDO. V balíčku AIP uloženém v celku Archivní úložiště je objekt CDO, který je replikou (bitově identickou odvozeninou) objektu CDO v balíčku SIP uloženém v dočasném úložišti funkčního celku Příjem. V archivu jsou dále balíčky AIP replikovány do jedné nebo více záloh. Archiv by měl označit, jaká z těchto odvozenin je směrodatná (tj. bude zdrojem vytváření balíčků DIP, případných budoucích formátových konverzí nebo obnovy ze zálohy apod.).

V archivu v procesu uchovávání lze teoreticky na objektu CDO v balíčku AIP vykonávat všechny typy transformací, které byly popsány v souvislosti s digitalizací (např. dodatečně ořezat obrazy), ale v praxi je relevantní především formátová konverze,

vykonávaná z důvodů zastarávání archivačního formátu digitálního originálu. K takovým konverzím však v současné praxi nedochází (archivační formáty TIFF a JP2 zatím nezastarávají). Dochází pouze ke konverzím z ekonomických důvodů (z formátu TIFF do formátu JP2 z důvodu zmenšení objemu uložených obrazových dat).

Formátové konverze musejí být vždy zapsány jako událost do provenienčních metadat. Metadata by však měla také odkazovat na dokumentaci celého procesu formátové konverze, popisující zejména důvod konverze, výběr nového archivačního formátu, proces konverze, užití nástroje i předchozí testování, na základě kterého byly vybrány tyto nástroje a způsob technické specifikace klíčových vlastností (tj. určení typové digitální identity) a mechanismus kontroly zachování těchto vlastností po skončení konverze. Objekt CDO první verze balíčku AIP by měl být vždy zachován, bez ohledu na to, zda jej bude možné zobrazit v budoucím technologickém prostředí (i v případě, kdy jej již nepůjde zobrazit, může sloužit alespoň jako částečný doklad autenticity). Zdůvodnění formátové konverze je důležité (např. konverze do ztrátové komprese z důvodu snížení nákladů na úložné kapacity je pochopitelné řešení).

#### **7.4.5 Odvozování při vytváření balíčku DIP**

##### **7.4.5.1 Optimální stav odvozování podle modelu OAIS**

V současné praxi je potřeba digitální knihovnu považovat za součást archivu (jako funkční celek Zpřístupnění), pokud jde o řízení celého životního cyklu digitalizátů jednou organizací. Skutečnost, že jsou užívané různé softwarové systémy pro různé funkční celky archivu OAIS (např. jeden pro celek Archivní úložiště, jiný pro celek Zpřístupnění), není v principu podstatná.

V optimálním případě by měl mít zvolen následující postup. Do digitální knihovny (jako součásti archivu) je dodán balíček AIP. Digitální knihovna převede obrazovou komponentu (jež je v archivačním formátu) objektu CDO tohoto balíčku do prezentačního formátu (meziformátu). Tato formátová konverze rovněž musí zohledňovat klíčové vlastnosti, ale nemusí zohledňovat potřeby digitální archivace. Prezentační formát musí být volen z hlediska potřeb čtenářů, zejména z hlediska jeho podpory v běžných internetových prohlížečích. Textová a strukturální komponenta musejí být rovněž zachovány v souladu s klíčovými vlastnostmi (tj. prezentovaný digitalizát musí dodržovat posloupnost stran tištěné předlohy a musí být prohledatelný). Výsledkem těchto odvozování je objekt CDO, který bude uložen v úložišti digitální knihovny. Následné otázky reprodukce této odvozeniny se věnuje oddíl 7.6.

#### **7.4.5.2 Současné problémy odvozování balíčku DIP**

Pokud se jako součást digitálního originálu vytváří v průběhu digitalizace také prezentační obrazová varianta a vše se v podobě balíčku SIP následně dodá do archivu, lze tento přístup považovat za přijatelný (nikoliv optimální) za předpokladu, že je celý řetězec odvozování prezentační varianty zdokumentován. Digitální knihovna vyjme z balíčku AIP pouze obrazovou komponentu v prezentačním formátu (meziformátu), což je proces replikace, jehož integritu stačí ověřit kontrolou neporušenosti na základě digitálního otisku.

Pokud jsou balíčky SIP dodávány současně do archivu a digitální knihovny (např. v projektu NDK), pak to znamená, že digitální knihovna není řízena jako součást archivu OAIS, ale jako odlišný systém. Vzhledem k tomu, že objekty CDO v odlišných formátech jsou odvozeny ze stejného zdroje a vytváří je tatáž organizace, může se na první pohled zdát takový přístup v pořádku. Problém však nastává v případě, kdy je potřeba vykonat opravy objektu CDO (v archivu, v digitální knihovně, nebo v obou systémech). Tento problém se žel v praxi vyskytuje běžně a jako o jednom z hlavních aktuálních rizik o něm blíže pojednává oddíl 7.7.

### **7.5 Problémy procesů transformace objektu CDO**

Problémy transformace objektu CDO lze rozdělit do tří fází: stanovení cílů transformace, volba transformačního nástroje a ověření výsledků transformace (validace).

#### **7.5.1 Stanovení cílů transformace**

Pro každou transformaci objektu CDO musejí být předem stanoveny cíle, kterých má být transformací dosaženo. Tyto cíle musejí zahrnovat otázky formátu a obrazových vlastností a musejí být v souladu s klíčovými vlastnostmi intelektuální entity. Stanovení těchto cílů odpovídá stanovení digitální identity nové generace objektu CDO.

Otázky formátu a formátového profilu byly již zmíněny v oddíle 7.3. Specifickou otázkou je stanovení obrazových vlastností v souladu s klíčovými vlastnostmi. Například pro klíčovou vlastnost „plná barevnost“ je nutno užít bitovou hloubku minimálně 24 bitů; pro vlastnost „dobrá čitelnost“ je nutno zvolit dostatečné prostorové rozlišení. Celkově lze říci, že v oblasti digitalizace existuje dostatek vhodných doporučení ve směrnících důvěryhodných organizací, přičemž tato disertační práce doporučuje řídit se směrnicí americké iniciativy FADGI. V případě pozdějších transformací (v archivu a digitální knihovně) je nutné stanovit

takové cíle transformace, které budou zachovávat klíčové vlastnosti digitálního originálu v souladu s aktuálním stavem technologického prostředí a požadavky archivace, resp. čtenářů.

### **7.5.2 Volba transformačních nástrojů**

Transformace objektu CDO mohou realizovat pouze digitální technologie, v některých případech je však nutná také manuální práce (např. ořezy). Transformační nástroje determinují možnosti výsledných vlastností nového objektu CDO. Tyto nástroje můžeme rozdělit nejméně do tří skupin: snímací zařízení, konverzní nástroje (nástroje pro formátovou konverzi) a nástroje pro intraformátové transformace (tj. v rámci téhož formátu, v praxi jde zejména o grafické aplikace). Před návrhem transformací by měl být proveden průzkum vhodných nástrojů, včetně studia odborné literatury, aby organizace nenavrhovala postup transformací, které nemůže realizovat (např. z finančních důvodů).

O snímacích zařízeních bylo již pojednáno v oddíle 7.4.2. Konverzní nástroje a nástroje pro intraformátovou transformaci mohou být v praxi funkcemi stejné aplikace (například software Adobe Photoshop umožňuje provádět obrazové úpravy i konverzi do jiných formátů).

Kvalitní konverzní nástroje pro vytváření formátu TIFF jsou široce rozšířené a řada z nich je nekomerčních. Pro konverzi do formátu JP2 však existuje mnohem menší počet nástrojů. Dlouhou dobu byly za kvalitní považovány pouze komerční produkty, zejména kodek firmy Kakadu, který je dosud patrně nejužívanějším nástrojem pro konverzi do formátu JP2. V poslední době se slibně vyvíjí opensourcový nástroj Openjpeg, jehož kvalita se oproti minulosti výrazně zlepšila. Kakadu ani Openjpeg zatím nedokáží vytvářet formát JP2 z formátů rodiny RAW. Tyto formáty je nutno nejprve jiným nástrojem převést do konverzního meziformátu TIFF, se kterým již dokáží tyto nástroje pracovat. Současná praxe transformace do formátu JP2 je tak výrazně omezena na užití uvedených nástrojů, což může představovat jisté riziko, pokud přestanou být dostupné nebo jejich vývoj bude ukončen. Užití nástroje Kakadu navíc představuje finanční náklady, které mohou zejména pro menší knihovny znamenat nepřekonatelnou překážku. To je případ většiny českých knihoven, které užívají opensourcový Openjpeg. Starší verze tohoto nástroje však nedokázaly vytvářet formát JP2 podle všech parametrů některých profilů, proto je třeba vysoká obezřetnost v tom, jakou verzi si organizace opatří. Je také nutné sledovat odborné studie posuzující kvalitu aktuálně dostupných nástrojů, zejména pro formátovou konverzi do JP2.

V oblasti grafických aplikací je nabídka široká, včetně nekomerčních nástrojů. V případě požadavku provést přiřazení barevného profilu Adobe RGB 1998 je otázkou,

nakolik kvalitně dokáží jiné aplikace než původce tohoto profilu provést takovou transformaci, a především, zda jiné aplikace nenabízejí svoji vlastní verzi ICC profilu pro převod do barevného prostoru Adobe RGB 1998. Dalším problémem je pak cena aplikací firmy Adobe.

### 7.5.3 Validace výstupů transformace

Ověření výsledků transformace musí být možné provést automatizovaně na všech objektech CDO. V některých případech je však nezbytná současná manuální kontrola. Validace by měla probíhat po každé transformaci, která vytváří novou generaci souborů. Jak již bylo uvedeno, některé dílčí transformace lze zahrnout do jedné komplexní transformace, a validace tedy musí probíhat až nad generací dat vzniklých po poslední dílčí transformaci.

Klíčovou otázkou je validace formátu. Definice validace se v praxi odvíjí podle definice uplatňované daným validátorem. V současnosti nejužívanějším univerzálním validátorem formátu je JHOVE.<sup>82</sup> Tento nástroj definuje tři úrovně validnosti (94): správná strukturovanost (*well-formedness*) znamená splnění syntaktických požadavků formátu, validita (*validity*) splnění dodatečných sémantických požadavků a konzistentnost (*consistency*) soulad s externími požadavky (94). Konzistenci lze označit za soulad s formátovým profilem. JHOVE ve své dokumentaci popisuje požadavky pro první dvě úrovně validnosti pro jednotlivé formáty, přičemž u řady formátů stanovuje, že požadavkem druhé úrovně validnosti je pouze to, aby formát splňoval požadavky úrovně první (tak je tomu v případě formátu JP2, zatímco pro formát TIFF je pro druhou úroveň validnosti stanoveno několik dostatečných požadavků). Nástroj jpylyzer, který je určen pouze pro formát JP2, odlišuje pouze jednu úroveň validnosti (JP2 je, nebo není validní).

Prvním krokem při validaci formátu však musí být formátová identifikace. V případě, že je identita požadovaného formátu stanovena na základě identifikátoru PUID, pak musí být užít nástroj pro formátovou identifikaci, který dokáže jako výstup identifikace vydat tento specifický identifikátor. Klíčovým nástrojem v této oblasti je DROID, který pochází od stejného původce, jenž provozuje registr PRONOM. Součástí registru je i identifikační mechanismus (způsob, jakým v souboru daného formátu zjistit jeho vlastní interní identifikátor, podle kterého lze určit formát a jeho verzi). Formátová identifikace pak v tomto

---

<sup>82</sup> Ve smyslu schopnosti validovat více formátů různých typů dat, nikoliv jako validátor všech existujících formátů. Takový validátor formátu neexistuje.

případě znamená, že nástroj DROID na základě identifikačního mechanismu zjistí typ formátu a jako výstup podá údaj o identifikátoru PUID (identifikátor PUID je externí identifikátor, není obsažen v souboru daného formátu). V poslední době se objevily další nástroje pro formátovou identifikaci, které také využívají identifikátor PUID a identifikační mechanismus registru PRONOM (Siegfried, FIDO). Jejich užití by však mělo být doplňkové vzhledem k jejich novosti a také tomu, že nepocházejí přímo od provozovatele registru PRONOM.

Následná validace formátovým validátorem je proces, kterým se zjišťuje soulad formátu daného souboru s jeho oficiální dokumentací (formátovou specifikací). Z hlediska digitální identity lze říci, že formát je autentický, pokud je v souladu se svou specifikací, tedy že je tím, za co se vydává. Z formátových validátorů jsou klíčové zmíněné nástroje JHOVE (pro TIFF, JP2 a XML) a jpylyzer (pro JP2), které by měly být užity vždy. Doplňkově je vhodné užít další formátové validátory.

Pro validaci struktury objektu CDO a strukturální komponenty při digitalizaci je klíčová manuální kontrola, která musí být vykonávána ve fázi finalizace digitálního originálu. Současně by v této fázi měla také být provedena vizuální kontrola kvality obrazových komponent, a to jak z hlediska dodržení postupu rekonstrukce (např. kvalita ořezů), tak vizuální kvality.

Pro validaci formátového profilu a obrazových vlastností je nutno užít extraktory technických metadat, které dokáží ze souborů získat technické informace a tyto výstupy srovnat se stanovenými cíli transformace. Funkce těchto extraktorů rovněž vykonávají nástroje JHOVE (pro TIFF, JP2 a XML) a jpylyzer (pro JP2), které jsou i v této oblasti vhodnými nástroji.

Nástroje JHOVE, DROID a jpylyzer jsou široce rozšířené a užívané delší dobu a lze je považovat za důvěryhodné. Jsou také opensourcovými nástroji a jsou dostupné zdarma. Jejich využití musí být základem softwarového řešení digitalizačního systému i archivačního systému, a to i v případě, že tyto systémy jsou komerční řešení. Spoléhat se na funkce komerčních nástrojů, které k identifikaci, validaci nebo extrakci mohou vytvořit komerční firmy, je postup, který nelze v žádném případě doporučit. Užití alternativních nástrojů může být vhodné, ale jen jako doplněk k těmto základním nástrojům. Pro automatickou validaci celého objektu CDO je nutné, aby si organizace vytvořila vlastní nástroj (komplexní validátor), který umožňuje přebírat výstupy uvedených formátových validátorů, extraktorů metadat a nástrojů pro formátovou identifikaci a srovnávat je se zadanými cíli transformace.



Problémem těchto specializovaných nástrojů je, že jejich novější verze mohou přinášet odlišné výsledky. V rámci přípravy digitalizace by proto měla být otestována jejich schopnost provádět požadované operace. Větší problém představují různé výstupy jiných verzí pro archivaci. Archivy Nového Zélandu například zaznamenaly případy, kdy nástroj DROID v nových verzích přináší odlišné výsledky (jiný identifikátor PUID), a zvažují provedení opětovné formátové identifikace svých digitálních dokumentů uložených v různých formátech (95).

Současná doporučení pro volbu archivačních formátů jsou založena na atributu otevřené dokumentace (dostupná formátová specifikace). Toto doporučení je navíc reflexí skutečnosti, že pokud formátová specifikace není dostupná, není možná transparentní validace formátu (může existovat pouze validátor původce, který jako jediný zná formátovou specifikaci, ale tento validátor z principu nemůže být transparentní). Rozvedením této teze je, že užití formátu s otevřenou dokumentací, pro který ale neexistuje validátor, představuje také jisté riziko. Problém je však i s existujícími zavedenými validátory, které se vyvíjejí a jejich nové verze také mohou přinášet odlišné výsledky. V současnosti probíhající evropský projekt PREFORMA<sup>83</sup> vyvíjí specializované opensourcové validátory pro formáty PDF/A a TIFF. To ukazuje na to, že současné validátory těchto formátů nejsou považovány za optimální.

Specifickou otázkou je validace obrazových vlastností, která je nad rámec možností formátových validátorů. V praxi projektu NDK se vyskytly případy souborů ve formátu JP2, které byly označeny validátory JHOVE i jpylyzer za validní, ale nebylo možné je zobrazit v digitální knihovně (41). Pro kontrolu obrazových vlastností je nutné užít speciální nástroje. V NK ČR jsou za tímto účelem užívány opensourcový nástroj ImageMagick<sup>84</sup> a komerční software firmy Kakadu. Oba vykonávají stejné funkce, ale jejich společné užití je odůvodněno zvýšením jistoty obrazové validnosti. Tyto dva nástroje jsou součástí komplexního validátoru, který v roce 2016 vyvinula jako opensourcový nástroj sama NK ČR (70). Tento komplexní validátor obsahuje také nástroje JHOVE a jpylyzer a umožňuje automatickou validaci formátového profilu a kontrolu metadat v balíčku SIP.

#### **7.5.4 Možnosti ověření věrnosti digitálního originálu k tištěné předloze**

Otázka věrnosti, jak bylo uvedeno, je otázkou stanovení referenčního objektu. Ten zahrnuje otázku tištěné předlohy a jejího časového určení, stanovení typové digitální identity

---

<sup>83</sup> <http://www.preforma-project.eu/>

<sup>84</sup> <https://www.imagemagick.org/>

digitálního originálu a klíčových vlastností a vztah digitálního originálu k předloze (z hlediska převáděné úrovně abstrakce podle FRBR a způsobu odvozování). Z hlediska intelektuální entity (zobrazeného digitalizátu) je důležitá možnost posouzení věrnosti srovnáním s tištěnou předlohou, které nelze vykonat jinak než lidským pozorováním. V současné praxi je zohledňován aktuální stav tištěné knihy v době digitalizace, a ten tedy může být předmětem posuzování věrnosti digitalizace. Obrazové úpravy, které doporučují provádět směrnice FADGI, jsou zamýšleny jako způsob, jak původní snímek upravit tak, aby věrněji reprezentoval předlohu. V dokumentaci musí být uvedeno, zda je odvozování založeno na těchto úpravách, nebo že se neprovádějí žádné úpravy. Záleží pak na posouzení čtenáře, jaký z těchto dvou přístupů bude považovat za záruku vyšší autenticity. Z hlediska dlouhodobého uchovávání však platí, že tištěná předloha dozná v průběhu času nutné důsledky stárnutím, a tak digitalizát bude v budoucnosti již reprezentovat odlišný objekt, než který bude představovat budoucí stav jeho původní předlohy.

Specifickým případem úprav, se kterým se někdy v současné praxi setkáme a který bez srovnání s předlohou nemusí být patrný (pokud není zdokumentován), je remasterování. Remasterováním se zde rozumí „vylepšování“ obrazu takovým způsobem, aby text byl výraznější (zvýšení kontrastu) a pozadí bělejší (tj. aby neprosvítal text z druhé strany). Tento přístup může být považován za částečné narušení autenticity digitalizátu knihy. V každém případě musí být uveden v dokumentaci popisující způsoby odvozování.

#### **7.5.5 Znalost limitů a zkreslení transformací**

Každá transformace znamená určitou ztrátu informačního obsahu předešlého objektu. Transformace jsou limitovány možnostmi dostupných technologií, které se časem mohou vylepšovat, což platí zejména o snímacích zařízeních. Proto dokumentace omezení technologií může pomoci budoucímu porozumění vlastnostem digitalizátů (např. nižší obrazové kvality z důvodu nedokonalosti nebo drahé ceny technologií). Takováto dokumentace může být v budoucnosti prostředkem, který umožní čtenářům v době, kdy bude dostupná mnohem lepší a levnější technologie, lépe porozumět volbám minulosti.

Dále je třeba uvést, že až se tištěné předlohy rozpadnou z důvodu degradace, pak o nich možná budeme vědět jen tolik, kolik nám „sdělí“ jejich digitalizáty. Jednou může být neocenitelnou možností srovnat digitalizáty různě starých knih nebo různých stavů exemplářů (v době snímání) pro rekonstrukci toho, jaká bývala povaha tištěných dokumentů, jakkoliv to budou vodítka nedostatečná (např. zakřivení stran bude odkazovat k trojrozměrnosti knihy).

Problémem je také komplexita obrazových dat, která se týká především otázky různých barevných profilů. Komplexitu barevných profilů ostatně ukazuje sama existence Mezinárodního konsorcia pro barvy (International Color Consortium), která definuje způsob vytváření ICC profilů za účelem zachování barevné věrnosti obrazových dat napříč různými systémy v současnosti. O to obtížnější bude ji zachovat v budoucnosti, až současné barevné profily zastarají.

Jakýkoliv archivační formát přináší určitá rizika. I v případě, že archiv zvolí některý z doporučených archivačních formátů, v dokumentaci archivu by mělo být uvedeno, na základě kterých kritérií byl zvolen a jaká jsou jeho známá rizika. Například na některé problémy specifikace formátu JP2 upozornil v roce 2011 ve svém článku van der Knijff (35). Tento článek může být použit jako součást dokumentace možných rizik formátu JP2.

## **7.6 Otázky zpřístupňování**

Knihovny již delší dobu volí jako formu zpřístupňování digitalizátů čtenářům jejich prezentaci v digitální knihovně (tj. zobrazení digitalizované knihy v její struktuře a s přidruženým popisem), která je určena pro internetové prohlížeče. Autenticita je tak především otázkou toho, za co organizace digitalizát knihy ve své digitální knihovně vydává, tj. prezentuje (otázka identity intelektuální entity) a jakým způsobem je reprodukován (otázka integrity intelektuální entity).

### **7.6.1 Organizační řešení zpřístupňování**

Pro řízení celého životního cyklu jednou organizací je nutné trvat na tom, aby organizace spravovala svoji digitální knihovnu jako součást archivu OAIS; to znamená, aby digitální knihovna představovala funkční celek Zpřístupnění. Není podstatné, že se užívají různá softwarová a hardwarová řešení pro digitální knihovnu na jedné straně (jako funkční celek Zpřístupnění) a ostatní funkční celky archivu OAIS na straně druhé. Podstatné je, aby mohl být archiv OAIS (zahrnující digitální knihovnu) řízen jako jeden celek.

Z toho vyplývá požadavek, aby se balíčky DIP vytvářely z balíčků AIP uložených ve funkčním celku Archivní úložiště. Balíčky AIP musejí být z toho celku archivu replikovány do dočasného úložiště digitální knihovny. Úložiště digitální knihovny je dočasné úložiště v tom smyslu, že poskytuje pouze bitovou ochranu – balíčky DIP budou uchovávány pouze tak dlouho, dokud nebude potřeba vytvořit z balíčků AIP nové balíčky DIP (z důvodu zastarání formátu objektu CDO v balíčku AIP nebo z důvodu změn požadavků na prezentační

formáty). Tyto nové balíčky nahradí původní, které budou opět předmětem bitové ochrany. Pro zpřístupňování je tedy klíčové, aby odvození prezentační varianty z formátu objektu CDO v balíčku AIP proběhlo bezchybně a prezentační varianta zachovávala klíčové vlastnosti intelektuální entity, s přihlédnutím k potřebám čtenářů (nikoliv archivace). To znamená, že prezentační formát může být i formát zatížený patenty.

Balíčky DIP digitální knihovna reprodukuje čtenářům ve své webové vrstvě. Zpracování balíčků DIP (tj. uložených objektů) při této reprodukci může být různé. Může zahrnovat obrázkový server, který prezentační formáty vytváří až „za chodu“ z prezentačního meziformátu (např. z formátu JP2 do JPEG), nebo mohou být prezentovány digitalizáty v již vytvořených prezentačních formátech (např. JPEG) a doplňkových formátech (např. GIF pro náhledy). Způsoby zpřístupnění se od počátku vzniku digitálních knihoven změnily, ale podstata zůstává – zpřístupňovat digitalizáty knih čtenářům v podobě prezentace pro webový prohlížeč.

#### **7.6.2 Prezentace jako způsob řešení problémů reprodukce**

Digitální knihovna při prezentaci digitalizované knihy čtenářům užívá zpřístupňovací software (jako způsob řešení problému interpretačních informací). Zpřístupňovací software tvoří tzv. ostatní interpretační informace (viz oddíl 2.2.2.1.). Jak již bylo uvedeno, problém integrity intelektuální entity je především problémem správné reprodukce objektu CDO: různé softwarové aplikace mohou interpretovat tentýž objekt CDO jinak. Digitální knihovna musí tedy ručit za to, že reprodukce je adekvátní. To znamená, že musí přizpůsobovat svoji prezentaci digitalizovaných knih aktuálním prohlížečům v závislosti na jejich vývoji a také sledovat kvalitu užitých softwarových komponent digitální knihovny (zejména zpřístupňovacího softwaru) a jejich případné zastarávání.

Požadavky cílové komunity na podobu a formáty zpřístupnění by měla knihovna zjišťovat také výzkumem mezi čtenáři. To není obvyklý jev v české knihovnické praxi. Tyto požadavky však lze v obecné rovině odvodit také z každodenní praxe: nepochybně nejlepším řešením je, aby obrazová data byla zobrazitelná přímo v běžném internetovém prohlížeči, bez nutnosti dodatečně instalovat různé pluginy. Pokud toho nelze docílit, musí organizace alespoň upozornit na potřebu takového pluginu a optimálně uvádět doporučení a instrukce, jaký plugin je vhodný, případně rizika, která se budou objevovat v průběhu času. Například NK ČR užívá dvě digitální knihovny pro novodobé fondy, což je dáno historickým vývojem. Stará digitální knihovna (Kramérius 3) nabízí jako prezentační formát DjVu. Tento formát byl kdysi vážným konkurentem formátu PDF, ale již výrazně zastarává a jeho užití je značně

omezené. Pro NK ČR je časově náročné převést obsah starého systému do nového. Proto zatím její formátový specialista testuje nové verze prohlížečů a informuje čtenáře na stránkách této digitální knihovny na případné problémy, viz: „Upozorňujeme uživatele digitální knihovny Kramerius [myšleno Kramerius 3, pozn. autora] na to, že s novou verzí prohlížeče Google Chrome mohou nastat problémy se zobrazováním dokumentů. Doporučujeme proto používat prohlížeč Internet Explorer nebo Mozilla Firefox, případně starší verzi prohlížeče Google Chrome.“ (96).

Digitální knihovna tedy sama o sobě (jako software) představuje interpretační informace. Aby takto mohla fungovat, musí průběžně sledovat vývoj jednak v oblasti zpřístupňovacích softwarových aplikací (např. softwaru od firmy Kakadu pro zpracování meziformátu v JP2 do prezentačního formátu JPEG), jednak v oblasti nejrozšířenějších internetových prohlížečů, kterým musí být prezentace digitální knihovny vždy přizpůsobena.

### **7.6.3 Možnost stažení jako doplňková forma zpřístupnění**

Možnost stažení celého digitalizátu je další formou zpřístupnění v digitálních knihovnách, ne však častou. V některých případech je nabízena možnost stáhnout celý digitalizát v jednom souboru v PDF, někdy se tento soubor generuje na vyžádání. Například systém Kramerius, užívaný ve většině digitálních knihoven v ČR, umožňuje generování PDF po dvaceti stranách.

Možnost, aby si čtenář stáhl celý digitalizát v jednom souboru, se zdá být základním požadavkem cílové komunity čtenářů (jeden soubor je vhodnější pro následnou práci a jednodušší na stažení). Digitální knihovny se však stále soustředí na zpřístupnění formou prezentace obrázků v prohlížeči. To je užitečný způsob zpřístupňování pro prohlížení různých knih, který však není nejvhodnější pro samotnou četbu celé knihy, zejména v kontextu současných čteček.

Pokud digitální knihovna nabízí ke stažení takový soubor v PDF, měl by zachovávat klíčové vlastnosti intelektuální entity (např. prohledatelnost). Optimálně by měl nabízet stejnou obrazovou kvalitu jako prezentované jednotlivé obrazy stránek. Pokud tomu tak není, musí být součástí souboru PDF vložená informace (optimálně na začátku textu) o tom, že jde o prezentační variantu v nižší kvalitě, než kterou mají prezentované obrazy v digitální knihovně. V každém případě by měla digitální knihovna uvést na svých stránkách přesnou specifikaci podoby stahovaných dat (např. formát PDF a jeho verzi) a instrukce, jakým způsobem (softwarem) doporučuje digitalizát prohlížet. Optimálně by měla též nabízet digitální otisk souboru, aby si čtenář mohl ověřit neporušenost souboru po jeho stažení.

Čtenáři by však měli mít také možnost získání balíčku AIP, a tedy se dostat k nejvyšší obrazové kvalitě digitalizátu, zejména pokud jde o digitální knihovny organizací, které digitalizují z veřejných zdrojů. Je zřejmé, že taková možnost bude náročná na softwarové řešení archivu, nicméně měla by být nabízena jako volitelná možnost. Jedním z možných řešení je například online objednávkový formulář, kterým si čtenář balíček objedná a archiv mu pak pošle oznámení emailem, až budou data připravena ke stažení. Archiv také může nabízet vytvoření speciálních uživatelských účtů pro stanovou skupinu čtenářů (např. vědecké pracovníky).

#### **7.6.4 Popisné informace prezentované spolu s digitalizátem**

Jako součást prezentace v digitální knihovně musejí být přidružené údaje o identifikačních informacích vztahujících se k digitalizátu knihy, které vyjadřují atributy jeho identity a jsou základem pro posuzování autenticity. Tyto vystavované informace musejí zahrnovat údaje o tištěné předloze na úrovni exempláře i vydání. Tyto údaje v principu většinou odpovídají údajům bibliografického záznamu. Dále musejí obsahovat údaje o digitalizátu a vztahu mezi digitalizátem a předlohou (minimálně informace o digitalizačním projektu, s odkazem na jeho dokumentaci). Bez těchto informací nemůže čtenář posoudit autenticitu digitalizátů a prezentovaný dokument lze hodnotit jako pochybného původu. Popisné informace musejí také jasně vyjadřovat, že jde o digitalizát knihy, nikoliv digitálně nativní dokument.

Všechny tyto informace musejí být obsaženy na úvodní stránce digitalizované knihy, ze které je možné se dostat na všechny části knihy (stránky) a případně na odkazy ke stažení celého dokumentu. Součástí konkrétní digitální identity objektu CDO v prezentačním formátu je URL adresa této úvodní stránky. URL adresa je součástí konkrétní digitální identity objektu CDO jakožto zpřístupňovaného objektu (balíčku DIP), nikoliv objektu CDO v balíčku AIP.

Pro čtenáře i pro dlouhodobou správu digitalizátů je však klíčové užití perzistentního identifikátoru, který není závislý na konkrétní lokaci, ale který současně dokáže na tuto lokaci kdykoliv přesměrovat. Tento identifikátor je nejdůležitějším prvkem pro otázky identity dokumentu v celosvětové internetové síti. Je také zárukou integrity citační praxe (jako řešení problému toho, že sice od okamžiku, kdy je digitalizovaná kniha přístupná v digitální knihovně, může být užívána a citována, přitom však mohla být změněna v důsledku opravy, resp. mohla být změněna URL adresa její úvodní stránky). Existence digitalizátu knihy v internetovém světě začíná až jeho zpřístupněním čtenářům. Aktuální lokace tohoto dokumentu v internetové síti je de facto hlavní prvek jeho konkrétní digitální identity

současného počítačového světa – prvek, který v kontextu fyzických dokumentů neexistuje, resp. nabývá absurdnosti.<sup>85</sup> Blíže se základním otázkám perzistentní identifikace věnuje oddíl 9.1.

#### **7.6.5 Autorizace digitální knihovny jakožto původce**

Na otázky autenticity zpřístupňovaného digitalizátu má také jistě vliv důvěryhodnost instituce. Užití certifikátů typu SSL je běžný mechanismus, jak zvýšit důvěru v identitu původce. Ideální je také udržovat doménový název i po jeho změně (jinak může být užít jinou organizací za jinými účely, což může být pro čtenáře matoucí, nebo může být zneužit). Smutnou praxí je, že například velká část evropských projektů, včetně těch, které se věnovaly digitální archivaci, již není dostupná na svých původních webových sídlech. Typickým příkladem je evropský projekt DigitalPreservationEurope, jehož jsem byl členem.<sup>86</sup>

### **7.7 Opravy digitalizátů knih**

Výše uvedené podkapitoly pojednávaly o standardním průběhu životního cyklu digitalizátů knih. Oprava znamená odchylku od navrženého (standardního) postupu. Opravy se mohou týkat nejen objektu CDO, ale i přidružených metadat. Z mé praktické zkušenosti vyplývá, že opravy (resp. chyby, které řeší) jsou jednoznačně hlavním rizikem pro udržení autenticity digitalizátů knih v současné praxi. Ve své odborné praxi jsem zaznamenal nejčastěji tyto druhy chyb:

- neúplné snímky (oprava: oskenování);
- nekvalitní snímky (oprava: opětovné skenování);
- nesoulad mezi obrazovými daty a přidruženými bibliografickými metadaty (bibliografický popis je popisem jiné knihy);
- strukturální změna dokumentu: rozdělení dokumentu na více dokumentů nebo spojení více dokumentů v jeden dokument

Z hlediska autenticity je nejproblematictější oprava balíčků DIP, tedy oprava realizovaná ve fázi, kdy jsou již digitalizované dokumenty prezentovány uživatelům a mohly být například citovány či jinak užívány. Takováto změna je z hlediska autenticity nepochybně

---

<sup>85</sup> Jak by mohl být údaj o aktuální geografické lokaci fyzické knihy součástí její identity? Absurdní požadavek pro analogový svět, ale klíčový pro digitální v kontextu internetové sítě.

<sup>86</sup> <http://digitalpreservationeurope.eu/>

největším rizikem (změny ve fázi před vytvořením balíčků DIP se koncového uživatele nedotýkají a jejich oprava je jen interní záležitostí organizace).

#### **7.7.1 Opravy balíčku SIP před jeho dodáváním do archivu**

Optimální situací je, aby si producent sám zkontroloval digitální originál a metadata v balíčku SIP. Pokud odhalí chybu u objektu CDO, je jednoznačně nejlepším řešením nová digitalizace. V optimálním případě totiž producent zaznamenává všechny změny objektu CDO do metadat, a pak by i oprava musela být nějak zaznamenána do metadat. To je jednak poměrně obtížné, ale především matoucí.

Producent by měl mít rovněž zvláštní postup pro opravy balíčků SIP, které po něm vyžaduje archiv. Pracovní postup pro opravy by měl být oddělen od běžné produkce. Pokud je to možné, opravy by měly být vykonávány dávkově. Pokud je pracovníků málo, pak je vhodné přerušit běžnou produkci a po nějaký čas se věnovat opravám chyb.

Producent by měl producent užívat validátory doporučené archivem, a to jak validátory dat (jejich formátů a obrazových vlastností), tak metadat. Pro metadata je taková validace poměrně komplikovaná, protože závisí na konkrétním aplikačním metadatovém profilu. Situaci dále komplikuje, pokud existuje více verzí takového profilu a různé projekty mají předepsané jiné metadatové profily. Komplikovanost takové validace ukázal v praxi vývoj komplexního validátoru v NK ČR (70).

Validací producent (např. daný organizační útvar jedné organizace) pomáhá především sám sobě, aby se vyhnul tomu, že bude muset dodávat balíčky SIP opětovně (na základě výzvy archivu k opravě chyby). Důležitá je transparentnost odpovědnosti za opravy zaznamenaná do dohody o dodávání dat (která může v případě stejné organizace mít čistou podobu jedné projektové dokumentace a odpovídajících popisů práce zaměstnanců různých útvarů). Producent a archiv musejí vědět, jaké nástroje na validaci užívají, a optimálně užívat stejné. Pokud archiv použije například novější verzi nástroje, než bylo uvedeno v dohodě o dodávání dat, de facto by neměl chtít po producentovi opravu chyb, které zjistil verzí nástroje, která je novější, než byla verze doporučená dohodou. V každém případě konečný verdikt o validaci je odpovědností archivu (validace uvedené v oddíle 7.5.3 jsou především otázkou archivu).

#### **7.7.2 Opravy při příjmu balíčku SIP do archivu**

Základním pravidlem by mělo být, aby archiv chybné balíčky SIP smazal a požadoval po producentovi dodání opravy. Může být totiž obtížné prokázat, že opravy provedené



archivem jsou skutečně opravami a nikoliv úpravami, které narušují autenticitu. V případě digitalizace v rámci téže organizace může jít o otázku odpovědnosti jednotlivých útvarů a také jejich pracovní náplně a odborné způsobilosti. Pokud opravy musí vykonat archiv, musejí se zaznamenat do metadat. Jistým řešením transparentnosti může být, aby archiv uložil chybný balíček SIP jako první verzi balíčku AIP, ze které by vzápětí vytvořil druhou verzi balíčku AIP (opravenou). Problém však může být s nastavením konkrétního technického systému. Například v NK ČR nelze uložit balíček AIP do archivního úložiště, pokud neprojde předepsanými kontrolami. V této fázi jde tedy o otázku toho, jakým způsobem řešit opravy v archivu, aby bylo možné doložit, že jde o opravy, nikoliv narušení autenticity. Proto je skutečně optimálním řešením, aby producent vytvořil nový balíček SIP (a to formou nové digitalizace, nikoliv oprav) a ten znovu zaslal do archivu.

### **7.7.3 Opravy balíčků AIP v archivu**

Opravy balíčku AIP mohou nastat za situace, že chyby nebyly zjištěny při příjmu balíčku SIP, ale zjistí se až po uložení balíčku AIP. U oprav balíčků AIP mohou nastat dvě situace: 1) z balíčku AIP již byl vytvořen balíček DIP a ten byl zpřístupněn (reprodukován) čtenářům; 2) čtenářům ještě nebyl zpřístupněn. Prvnímu problému se věnuje oddíl 7.7.4

V druhém případě je opět nejvhodnějším řešením vyžádat si od producenta nový balíček SIP. V tomto případě by však archiv neměl vytvářet novou verzi balíčku AIP, ale nový balíček AIP a původní balíček AIP smazat. Toto je možno i v případě, že to konkrétní technologie neumožňuje (viz ochrana pásek před smazáním). Pro tyto účely jsem navrhl koncept deaktivace dokumentu. Deaktivace dokumentu znamená, že balíček AIP (obsahující vadný dokument) bude označen jako deaktivovaný, což znamená, že při budoucí replikaci na nové datové nosiče nebude předmětem této replikace, čímž zanikne.

Je však také možné, že balíček SIP již nelze znovu dodat. Pak musí archiv sám provést opravu. Pokud opravu již vykonat nelze, musí archiv do metadat zaznamenat informace o tom, že dokument je porušený (např. chybí stránky).

### **7.7.4 Opravy balíčku DIP**

Tento typ opravy je z hlediska autenticity hlavní problém. Intelektuální entita již byla prezentována uživatelům, a tedy mohla být užívána (citována apod.). Oprava musí probíhat tak, že bude do dočasného úložiště funkčního celku Zpřístupnění replikován balíček AIP, ten bude opraven a znovu zaslán do funkčního celku Příjem. Zde bude převeden do nového balíčku AIP (původní bude smazán nebo deaktivován). Pak bude z tohoto nového balíčku AIP

vytvořen nový balíček DIP, který bude zobrazen v digitální knihovně. Původní balíček DIP musí být z digitální knihovny smazán. Tyto operace opět poukazují na nutnost pojímat digitální knihovnu jako součást archivu OAIS. Na stejné URL adrese by neměl být vystaven nový dokument.

## **7.8 Dokumentace a metadata**

Dokumentace a metadata jsou dodatečná data spojená s informačním obsahem, přičemž dokumentace je psána v přirozeném jazyce, zatímco metadata v umělém jazyce a jsou vysoce strukturovaná. Identifikátory lze chápat jako specifický typ metadat; speciálně perzistentním identifikátorům se věnuje Kapitola 9. Podle některých odhadů tvoří náklady na vytváření metadat až polovinu nákladů na řízení celého životního cyklu digitálního dokumentu (97). To může být ještě podceněným odhadem, pokud započítáme i vytváření dokumentace. Vytváření metadat i dokumentace vyžaduje práci specialisty. Digitální metadata ve své komplexitě zdaleka přesahují tradiční knihovnická metadata (katalogizační záznamy). Metadata a dokumentace by měly z hlediska autenticity zaznamenávat a uchovávat zejména atributy identity a integrity intelektuální entity (popis intelektuální entity a mechanismů jejího zobrazování) a atributy identity a integrity objektu CDO (popis typové a konkrétní digitální identity a řetězce odvozování objektu CDO).

### **7.8.1 Záznam atributů identity intelektuální entity**

Velmi důležitá jsou metadata a dokumentace zachycující popis referenčního objektu. Z hlediska metadat je nutno zaznamenat zejména údaje o exempláři a vydání tištěné knihy (včetně perzistentního identifikátoru vydání tištěné knihy), perzistentní identifikátor digitalizátu knih, datum vzniku digitalizátu a aktuální URL adresu umístění digitalizátu v digitální knihovně. Tato adresa se bude měnit, ale v daný okamžik musí vždy nějaká existovat jako aspekt internetové existence digitalizátu. Tento problém řeší užití takového systému identifikace, který dokáže pracovat s perzistentními identifikátory a současně s aktuálními adresami URL. Blíže o tomto problému pojednává Kapitola 9.

Dále je nutno zaznamenat klíčové vlastnosti intelektuální entity, tedy vlastnosti, které byly definovány v projektu a které se organizace zavazuje udržovat v průběhu času. Vzhledem k tomu, že metadatové schéma pro záznam signifikantních vlastností digitalizátů knih v současnosti neexistuje, je vhodným postupem zaznamenat tyto klíčové informace jednak do digitalizační dokumentace (v části popisující cíle digitalizace), jednak do dokumentace archivu (v části popisující závazek archivu).

Digitalizační dokumentace by dále měla obsahovat informace o způsobu odvozování z tištěné knihy do digitálního originálu, tedy všech transformací, a údaj o tom, na jaké úrovni abstrakce podle FRBR je kniha snímána.

Všechny výše uvedené informace musejí být současně prezentovány digitální knihovnou spolu s digitalizátem knihy (v podobě metadat nebo odkazů na dokumentaci). Tyto informace zde budou sloužit jako základ pro zjištění identity digitalizátu čtenářem, a tím možnost posouzení jeho autenticity.

### **7.8.2 Záznam atributů integrity intelektuální entity**

Integrita intelektuální entity se týká jejího správného reprodukování. V kontextu současné praxe jde tedy zejména o vhodně zvolený zobrazovací software a jeho dokumentaci. Archiv ručí za to, že zobrazovací software adekvátně zobrazuje digitalizované knihy v aktuálních verzích internetových prohlížečů. Měl by též uvést informace o všech prohlížečích, pro které byly provedeny testy zobrazení. Je pravděpodobné, že archiv nebude moci testovat všechny existující internetové prohlížeče, měl by však vybrat alespoň nejužívanější a jejich vhodnost pravidelně testovat. Záznamy tohoto testování pak průběžně doplňovat do digitální knihovny tak, aby bylo zřejmé, na jakou konkrétní verzi prohlížečů je digitální knihovna nastavena.

V případě, že archiv nabízí ke stažení celý digitalizát v jednom souboru (např. PDF), musí uvést informaci, jaká verze a typ PDF je užit a doporučit vhodné nástroje, resp. alespoň popsat ty, na kterých bylo ověřeno správné zobrazení. PDF by mělo obsahovat informace o identitě intelektuální entity a uvést všechna případná omezení míry věrnosti (např. PDF může být pouze černobílý text, zatímco digitalizáty prezentované v digitální knihovně zachovávají plnou barevnou věrnost).

### **7.8.3 Záznam atributů identity objektu CDO**

Atributy identity objektu CDO jsou rozhodné interpretační informace, důležité pro adekvátní reprodukci v aktuálním počítačovém prostředí a pro potřeby dlouhodobého uchovávání. V průběhu času se budou měnit, zejména v souvislosti s formátovou konverzí. V daný okamžik však musejí být aktuální atributy zaznamenány. K tomu slouží zejména technické informace, jejich konkrétní podobu určuje zejména standard MIX a odpovídající elementy standardu PREMIS. Mezi rozhodné interpretační informace o typové digitální identitě patří zejména informace o formátu, formátovém profilu, prostorovém rozlišení, bitové

hloubce, barevném modelu a profilu a kompresi; mezi rozhodné interpretační informace o konkrétní digitální identitě zejména pixelová velikost, velikost souboru a jeho název.

Informace o typové digitální identitě musejí být zaznamenány do digitalizační dokumentace (jako cíl digitalizace). Při všech transformacích musejí být voleny takové hodnoty těchto atributů, které jsou v souladu s klíčovými vlastnostmi intelektuální entity. Na úrovni digitální identity lze rovněž specifikovat tyto vlastnosti, ale tato specifikace se může časem měnit. Například lze stanovit, že klíčovou vlastností je barevný profil Adobe RGB 1998 a tuto vlastnost udržovat napříč formátovými konverzemi. Až tento profil zastará, bude nutné jako klíčovou vlastnost zvolit takový, který jej nahradí, ale jakým způsobem tento výběr bude probíhat, nevíme. Lze samozřejmě zapsat tuto klíčovou vlastnost jako „barevný profil Adobe RGB 1998 nebo podobně kvalitní“, ale takovou hodnotu pak nelze automaticky validovat. Proto je vhodnější v dokumentaci archivu uvést, že klíčové vlastnosti digitální identity budou v průběhu času upravovány.

#### **7.8.4 Záznam atributů integrity objektu CDO**

V zásadě platí, že všechny atributy identity objektu CDO se po jeho transformaci stávají součástí provenienčních informací, jako informace o předchozích generacích, které zpravidla již nebudou zachovávány. Provenienční informace, obsahující interpretační informace minulých objektů, jsou důležité z hlediska integrity objektu CDO v průběhu času a slouží jako prověřovací záznam. V kontextu digitalizace jako záznam pro ověření vytvoření digitálního originálu podle cílů digitalizace, v následujících fázích jako ověření vhodnosti formátových konverzí (v archivu nebo digitální knihovně).

Mezi informace o integritě objektu CDO, které jsou důležité z aktuálního hlediska, patří především informace o digitálním otisku a mechanismu a datum jeho vytvoření a informace provedených kontrolách neporušenosti souboru a validnosti formátu.

#### **7.8.5 Determinace metadatovými standardy**

Metadatové standardy jsou nutně determinující prvek. Řídit se mezinárodními (nebo široce užívanými) standardy znamená mnohem vyšší záruku interoperability. Zároveň to vyžaduje vyčkávat na případné doplnění elementů pro něco, co zatím popsat nelze, v budoucí nové verzi standardu či zcela novém standardu. V současnosti například neexistuje standardizovaný metadatový popis pro digitalizované zvukové dokumenty nebo pro klíčové vlastnosti (signifikantní vlastnosti PREMIS). Příkladem významné změny metadatového

standardu je vydání nové verze (č. 3) standardu PREMIS v roce 2015 (52). Až tato verze umožňuje zápis událostí na intelektuální entitě (například změnu bibliografických metadat).

Vytvářet vlastní metadatový standard je intelektuálně náročné a přináší riziko interoperability i nejistotu vývoje. Vytvářet externí schémata (jako rozšíření stávajících standardů) je smysluplnější. Z hlediska klíčových vlastností je zřejmě nejlepší stanovit klíčové vlastnosti intelektuální entity v dokumentaci, a tyto převést do podrobnějších elementů až v době, kdy budou dostupné ve standardu PREMIS.

Standard PREMIS také „vnucuje“ určitou logiku zápisu metadat. Změna objektu znamená vždy vytvoření nového objektu a nutnost zaznamenat ji jako vztah mezi zdrojovým a cílovým objektem (není možné zapsat verzi objektu). Tato logika je racionální a užitečná jako základní instrukce pro vytváření záznamů o provenienci pro knihovny, které nemají kapacity na to porozumět všem požadavkům životního cyklu z hlediska autenticity.

Dva neužívanější komerční archivační softwarové systémy (Preservica a Rosetta) užívají vlastní metadatová schémata, která jsou na vysoké úrovni komplexity (například schéma Rosetty umožňovalo zápis událostí na intelektuální entitě dříve než PREMIS). Problém však je, že jde o proprietární schémata. Je vhodnější trvat na užití standardu PREMIS a postupně aplikovat jeho novinky, než se spoléhat na komerční řešení svázaná s jednou firmou.

#### **7.8.6 Aplikační metadatový profil a způsob plnění metadatových elementů**

Metadatový profil užitý pro digitalizaci a metadatový profil užitý pro archivaci musejí být zdokumentovány (první v digitalizační dokumentaci, druhý v dokumentaci archivu). Metadatový profil musí obsahovat informace o všech využitých metadatových standardech a jejich verzích a musí být v souladu s těmito standardy. Měl by však také obsahovat vlastní řízené slovníky organizace, které by měly být užity zejména pro typologii událostí. Pro archivaci sice existuje například řízený slovník událostí pro PREMIS, který je ale příliš malý na to, aby dokázal popsat všechny typy transformací (pro digitalizaci navíc není původně určen).

Metadatový profil by však měl obsahovat i informace o tom, jakým způsobem mají být metadata plněna. To je základním požadavkem pro samu autenticitu metadat. Bibliografické údaje by měly být vždy získávány z katalogizačního systému, což znamená, že nejprve by měla být provedena jejich kontrola nebo rekatalogizace, a teprve pak transformace do standardu MODS. Velmi problematická, jak se ukázalo v praxi, je pozdější

oprava bibliografických metadat, protože ta vždy zasahuje identitu intelektuální entity prezentovaného digitalizátu, a tím ohrožuje autenticitu.

Interpretační informace (informace o typové a konkrétní digitální identitě) by měly být získávány přímo ze souborů užitím metadatových extraktorů v maximální možné míře. Postup získávání metadat by neměl být založen na údajích, které se přednastaví do digitalizačního systému (např. název skeneru pro jednu linku) a systém je následně automaticky přiděluje všem dokumentům dané linky. Dále je nutno, aby při snímání byla do původního snímku zabudována EXIF metadata. Tato metadata pak musejí být extrahována ihned z původních snímků, ne z následných generací. Existuje doporučení Kongresové knihovny, jaká EXIF metadata je minimálně třeba získat (98). Tento způsob získávání metadat je vzhledem k vysoké standardizaci EXIF možno označit za důvěryhodný způsob plnění metadat. Současně umožňuje předejít problému, kdy v lince dojde k náhradě snímacího zařízení, ale zapomene se tento údaj zadat do přednastavených hodnot.

Výstupy extraktorů však ne vždy lze jednoduše namapovat do metadatových schémat MIX a PREMIS. Metadatový profil musí také obsahovat způsob převodu informací z výstupů metadatových extraktorů do elementů metadatového profilu. Současně však může být vhodné, zejména ve fázi archivace, zaznamenat i samotná metadatová schémata těchto extraktorů a doplnit je jako externí schéma. Výstup nástroje FITS lze například začlenit jako externí schéma do PREMIS.<sup>87</sup>

### **7.8.7 Kvalita bibliografických záznamů**

Zjevným problémem je kvalita bibliografických záznamů a jejich kontrola před digitalizací. Optimálním způsobem je vytvářet co nejúplnější záznamy. Je potřebné co nejvíce využít potenciál bibliografického záznamu pro potřeby digitalizace, k čemuž by mohla sloužit katalogizační politika organizace. V této oblasti je klíčovou otázkou existence perzistentního identifikátoru na úrovni vydání pro všechny knihy (identifikátor ISBN v tomto ohledu nestačí). Blíže se tomuto problému věnuje Kapitola 9.

#### **7.8.7.1 Autenticita metadat**

Otázka autenticity metadat znamená, nakolik lze metadatům důvěřovat. Roli v tom nepochybně sehrává míra transparentnosti způsobu získávání metadat a související možnost

---

<sup>87</sup> Do sekce „objectCharacteristicsExtension“

ověření. Pro zvýšení redundance (a tím i vyššího zabezpečení metadat) je běžná praxe ukládat metadat jak v balíčcích AIP, tak v databázi.

Pro autenticitu je důležitá identita metadat, tj. přesná identifikace metadatového schématu, včetně verze, původce a zdroje. Zvýšením zabezpečení (pro potřeby budoucího ověření správnosti užití či interpretace) nepochybně bude uložení samotného metadatového standardu spolu s balíčky AIP do archivu. V budoucnosti totiž již tyto standardy nemusejí být dostupné online (běžným způsobem identifikace metadatového schématu je URL adresa). Takto uložený standard může být také dokladem toho, co bylo skutečně použito při vytváření metadatového profilu. Zvýšená obezřetnost je nutná u aplikačních metadatových profilů, která si organizace vytvořila pro svoje potřeby. Jejich správná identifikace a uchovávání všech verzí je nezbytností.

### **7.8.8 Dokumentace standardů a nástrojů**

Jistým problémem je úroveň dokumentace dostupných nástrojů a standardů. Metadatové standardy udržované Kongresovou knihovnou jsou poměrně příkladně uchovávány, včetně všech verzí. Některé standardy není možno získat jinak než koupí, což je zejména případ standardů ISO. V případě nástrojů je problém, že jsou často přesouvány na jiná webová sídla pro softwarové projekty. To potkalo řadu softwarových projektů na Google Projects, které nyní přesídlily na GitHub. Z minulosti si lze vzpomenout na Sourceforge.net, který sice funguje dodnes, ale již se stal nedůvěryhodným serverem.

Problém je i úroveň dokumentace nástrojů. Například JHOVE, jeden z hlavních nástrojů digitální archivace, na dvou odlišných stránkách jednoho ze svých sídel uvádí odlišné definice validnosti.<sup>88</sup> To ukazuje na vývoj dokumentace, ale působí to značně rozpačitě, když jde o definice tak klíčové části. Nakonec i samotné sborníky klíčové konference oboru digitální archivace (iPRES) se již několikrát přesunuly k jiným.

Řada nástrojů vyvinutých minulými evropskými projekty již není podporována, dokonce ani dostupná, jde zejména o nástroj PLATO pro plánování uchovávání. V současnosti jsou velké naděje vkládány do evropské iniciativy Open Preservation Foundation, na jejichž webovém sídle nalezneme řadu nástrojů i jejich dokumentaci (zde je i nástroj JHOVE, který je tedy dostupný na více místech, což může být matoucí).

---

<sup>88</sup> Na hlavní stránce jsou pouze dva typy (119), na jiné stránce téhož webu tři (120).

Organizace by měla v optimálním případě obstarat veškerou dokumentaci a uložit ji spolu s užitými verzemi speciálních nástrojů (DROID apod.) do archivu. Tato dokumentace by měla zahrnovat dokumentaci všech užitých metadatových standardů, nástrojů a formátové specifikace archivačních formátů. To může někdy vyžadovat finanční zdroje (formátovou specifikaci pro JP2 nebo normu ISO 14721 je nutno zakoupit od organizace ISO). Než užívat návrhy norem volně dostupné na internetu je bezpečnější opatřit si oficiální vydání norem od normalizačních organizací. Ideálním požadavkem je mít všechny standardy rovněž vytištěné.

## **7.9 Shrnutí**

Z přehledu problémů autenticity digitalizátů knih lze za rozhodné oblasti autenticity považovat následující kategorie.

### **7.9.1 Proceduralita**

Zcela zásadní je dodržení procedurálního principu, který je navržen normou ISO 14721 pro odvozování informačních balíčků (posloupnost SIP – AIP – DIP). Specifikum normy ISO 14721 spočívá v tom, že objekty CDO jsou nejprve uloženy do archivního úložiště archivu OAIS v balíčcích AIP, a teprve poté jsou z balíčků AIP vytvářeny balíčky DIP pro zpřístupňování. Tato proceduralita je, podle empirického zjištění autora této práce, klíčová zejména pro odvozování balíčků DIP a potřebu jednotné správy uchovávání a zpřístupňování. Digitální knihovnu je v kontextu řízení celého životního cyklu jednou organizací nutno pojímat jako součást archivu OAIS, nikoliv jako oddělený systém. Dodržování tohoto principu není důležité z nějakého čistě formálního hlediska, ale proto, že tento princip podporuje požadavky autenticity. V případě jakýchkoliv změn musí být nejprve změněn balíček AIP, a poté z něj musí být vytvořen nový balíček DIP v digitální knihovně. Znamená to tedy jednotné a centrální řízení změn (což jsou v současné praxi zejména opravy), a s tím související jednotné řízení perzistentních identifikátorů a synchronicitu mezi uchovávaným a zpřístupňovaným objektem. Není možné provádět změny v digitální knihovně, tím dochází k narušení vazby mezi oběma objekty.

### **7.9.2 Identita a identifikace**

Identifikace obecně je základním prvkem zjištění identity digitalizátů knih (jako jedné ze složek autenticity). Identitu je nutno v případě digitalizátů knih chápat na dvou úrovních. Jednak jde o identitu intelektuální entity, která zahrnuje identitu digitalizátu jako zobrazené



(vnímatelné) jednotky intelektuálního obsahu, identitu předlohy jako zdroje digitalizátu a vztah mezi digitalizátem a předlohou. Jednak o digitální identitu, která specifikuje konkrétní parametry aktuálního objektu CDO, jehož zobrazením intelektuální entita vzniká.

Z hlediska identity intelektuální entity je důležitá existence perzistentního identifikátoru digitalizátu, který dokáže odlišit různé digitalizace různých organizací v různém čase, jejichž předmětem může být tatáž kniha. Z hlediska identity předlohy je nejdůležitějším prvkem perzistentní identifikátor vydání. Vztah mezi digitalizátem a předlohou musí být vyjádřen v digitalizační dokumentaci. Digitální identita je důležitá jako stanovení cíle každé transformace objektu CDO, v digitalizaci jde o specifikaci výstupu digitalizace (digitálního originálu), při archivaci a zpřístupnění jde zejména o specifikaci výstupu formátové konverze, a to buď z hlediska potřeb uchovávání, nebo z hlediska potřeb zpřístupňování cílové komunitě (tedy čtenářům digitální knihovny). Digitální identita se může v průběhu času měnit, ale identita intelektuální entity musí být zachována. Proto je důležitou součástí identity intelektuální entity stanovení jejích klíčových vlastností (např. plná barevnost, plnotextová prohledatelnost), které musejí být zohledněny při stanovení cílů všech transformací.

Digitální identita je specifikem digitálních dokumentů. U fyzických intelektuálních entit (to, co vidíme např. při prohlížení knihy) není při uchovávání zpravidla potřeba zjišťovat a zaznamenávat všechny atributy fyzické identity knihy jakožto fyzického nosiče (např. váhu, chemické složení papíru), v běžné praxi by takový požadavek byl absurdní. V případě potřeby lze atributy fyzické identity zjistit například forenzní analýzou a tyto prvky jsou u tištěných knih pevně svázány s informačním obsahem; nelze je spravovat odděleně jako v případě digitálních dokumentů. V digitálním světě, z hlediska kvality produkce, správnosti konverzí a možností ověření validnosti objektu CDO je nutné digitální identitu vždy specifikovat. Různé aplikace zacházejí s objekty CDO různými způsoby (viz kvalita formátové konverze nebo reprodukce). Nejzřetelnějším požadavkem je autenticita formátu – aby byl formát tím, za co se vydává, tj. v souladu s jeho oficiální specifikací. Pro autenticitu formátu je důležité přesně stanovit jeho identitu (což umožňuje jen identifikátor PUID) a poté ověřit, zda soubor v daném formátu je validní, tedy jeho integrita odpovídá formátové specifikaci.

### **7.9.3 Validace jako kontrola integrity**

Validace integrity digitalizátu je formou technické kontroly jeho integrity a jde o komplexní proces. Validace integrity objektu CDO se sestává z celku validací integrity jeho digitálních komponent (obrazové, textové, strukturální komponenty). Zahrnuje kontrolu neporušenosti, validaci formátu nebo kontrolu požadovaných obrazových vlastností

a vyžaduje užití specializovaných nástrojů a komplexního validátoru, který dokáže tyto nástroje integrovat. Validaci integrity zobrazeného objektu CDO (tedy intelektuální entity) musí archiv provádět tak, že průběžně testuje kvalitu zpřístupňovacího softwaru, zda jeho zobrazovací schopnost odpovídá klíčovým vlastnostem intelektuální entity nebo nevykazuje chyby.

Zatímco validace integrity informačního obsahu jako uloženého objektu je primárně technologická záležitost, validace integrity intelektuální entity (reprodukováného informačního obsahu) se neobejde bez lidské kontroly. Problémem je, že různé softwary mohou reprodukovat intelektuální entitu různě, často s nepatrnými odchylkami. V případě validací digitálními technologiemi pak může být problémem různá kvalita užitých nástrojů, odlišnosti výstupů jejich různých verzí, ale také jejich dostupnost, udržitelnost a adekvátní dokumentace.

#### **7.9.4 Dostupnost a důvěryhodnost specializovaných nástrojů**

Bez specializovaných nástrojů nelze provádět digitální archivaci, a tedy ani posuzovat autenticitu. Přesto je velkým problémem jejich důvěryhodnost, daná nejen důvěryhodností organizace, ale také mírou kompletnosti dokumentace a udržitelností nástroje. Všechny slibné nástroje, vyvinuté velký evropským projektem PLANETS, jsou již buď nedostupné, nebo se jejich vývoj zastavil. Aktuálně slibným evropským projektem je iniciativa Open Preservation Foundation. Důležité je, aby zůstal udržován a vyvíjen registr PRONOM, jeho nástroj DROID a jeho identifikační mechanismus formátu. V americkém prostředí patří mezi důležité aktivity činnost Harvardovy univerzity (nástroje JHOVE a FITS).

Problémem je i sám vývoj nástrojů, které v nové verzi přinášejí jiné výsledky. To je sice typický jev jakéhokoliv softwaru, ale pokud jde o identifikaci a validaci formátů, přináší to značná rizika. Proto je optimálním řešením opakovaná identifikace a validace již uložených objektů CDO v balíčcích AIP. To je však časově a technicky nesmírně náročná operace. Může se však odhalit chyba, kterou lze ještě opravit; v budoucnosti už taková náprava nemusí být možná, a jak se bude taková situace řešit, zůstává noční můrou digitálních archivů.

#### **7.9.5 Úplnost a rozsah dokumentace**

Digitalizační dokumentace představuje největší riziko v oblasti dokumentace, v tom smyslu, že popisuje projekt, který skončí, a pokud nejsou zaznamenány všechny jeho volby a technické specifikace včas, později již nemusí být možné je dohledat. Totéž platí o předmětu metadat zaznamenávaných v digitalizaci (zaniklé starší generace dat, informace

o užitém skeneru apod.). Archiv by měl veškerou digitalizační dokumentaci shromáždit a uchovávat spolu s objekty CDO. Optimální je shromáždění a uchovávání všech užitých standardů a nástrojů jak v digitalizaci, tak při archivaci. Při zpřístupnění takový požadavek není zásadní, neboť požadavky zpřístupnění se odvíjejí podle aktuálních potřeb. Zde je však důležité zdokumentovat mechanismus perzistentního identifikátoru digitalizátu, aby byl znám čtenářům a mohli těžit z jeho výhod.

#### **7.9.6 Opravy jako aktuální problém**

Z praxe se ukazuje, že hlavní aktuální problém pro autenticitu není otázka formátové konverze v archivu, ale opravy chyb digitalizovaných dokumentů. Způsob provádění oprav úzce souvisí s proceduralitou. Hlavní problém představuje, pokud již v digitální knihovně byl prezentován digitalizovaný dokument, který je chybně vytvořený nebo došlo ke změně atributů jeho identity, a je potřeba jej opravit (nebo opravit metadata popisující tyto atributy). Je potřeba zachovat informace o původním (chybně vytvořeném nebo popsáném) dokumentu, a současně upozornit na změnu dokumentu. Dokument již mohl být citován či jinak užíván. Řešením tohoto problému může být specifická funkcionality systému perzistentní identifikace digitalizátů knih (viz oddíl 9.2.2).

#### **7.9.7 Komplexita obrazových dat**

Při rastrové digitalizaci je nejnáročnější operací zpracování obrazových dat. Užití ztrátové komprese pro archivní formát je neprozřetelné, a děje se jen kvůli finančním důvodům, ne z důvodu zastarávání formátu. Tytéž organizace, které jsou daleko s výzkumem digitální archivace, současně přistupují ke ztrátové kompresi (z důvodu ušetření na úložných kapacitách), což je jeden z paradoxů současné teorie a praxe digitální archivace. Specifickým problémem je otázka barevné specifikace. Ta je velmi složitá a ne vždy řešena. Například v projektu NDK není řešen problém ICC profilů a jejich převodu do zavedeného barevného profilu. Otázky komprese, barevné specifikace a kvality kodeků na datovou transformaci budou vždy představovat vysoké nároky na přípravu digitalizačního projektu i dlouhodobé uchovávání, a to nejen v souvislosti s komplexitou formátu JP2, ale také problémy validace, která například v souvislosti s formátem TIFF řeší aktuální evropský projekt PREFORMA.

## 8 Návrh postupu pro udržení autenticity digitalizátů knih

Tento doporučený postup pro udržování a posuzování autenticity digitalizátů knih v celém jejich životním cyklu je určen pro organizaci, která řídí sama celý cyklus. Postup zahrnuje obecná doporučení (sady požadavků „P-0“ a „P-1“) a doporučení pro jednotlivé fáze životního cyklu (sady požadavků „P-2“, „P-3“ a „P-4“).

### 8.1 Obecný procedurální postup (sada „P-0“)

P-0.1	Užít systém perzistentní identifikace vydání tištěných knih
P-0.2	Užít systém perzistentní identifikace digitalizátů knih
P-0.3	Řídit digitální knihovnu jako funkční celek Zpřístupnění
P-0.4	Stanovit principy vytváření balíčků SIP, AIP a DIP
P-0.5	Stanovit rozdělení odpovědností v rámci organizace

Požadavky sady „P-0“ musí splňovat každá organizace před zahájením jakékoliv digitalizace tištěných knih a týkají se nejen digitalizace, ale archivace a zpřístupňování.

#### **Užít systém perzistentní identifikace vydání tištěných knih (P-0.1)**

Tento požadavek znamená buď užít existující systém identifikace, nebo jej zavést a užít. Zavést takový systém však může jen organizace poskytující služby pro celý národní nebo mezinárodní kontext. Tento požadavek dále rozvádí sada požadavků „PID-K“, uvedená v oddíle 9.2.1.

#### **Užít systém perzistentní identifikace digitalizátů knih (P-0.2)**

Tento požadavek znamená buď užít existující systém identifikace, nebo jej zavést a užít. Zavést takový systém však může jen organizace poskytující služby pro celý národní nebo mezinárodní kontext. Tento požadavek dále rozvádí sada požadavků „PID-D“, uvedená v oddíle 9.2.2.

#### **Řídit digitální knihovnu jako funkční celek Zpřístupnění (P-0.3)**

Tento požadavek znamená, aby digitální knihovna by zavedena jako funkční celek Zpřístupnění archivu OAIS podle normy ISO 14721. Znamená to jednotné řízení archivace a zpřístupňování v souladu s touto normou (viz oddíl 2.2.3). Požadavek „P-0.3“ zahrnuje především nutnost vytvářet balíčky DIP až v digitální knihovně, a to takovým způsobem, aby objekt CDO v balíčku AIP (který je v archivačním formátu), zde byl převeden do objektu CDO v prezentačním formátu (meziformátu), což se týká zejména obrazové komponenty. Současně musí být zachována textová a strukturální komponenta, nebo její klíčové vlastnosti

(tedy zejména prohladatelnost a posloupnost stran) při prezentaci digitalizátu v digitální knihovně.

#### **Stanovit obecné principy vytváření balíčků SIP, AIP a DIP (P-0.4)**

Tento požadavek předpokládá dodržování modelu OAIS a znamená vytvoření zpřesňujících pravidel v rámci možností tohoto modelu. Postup vytváření balíčků SIP musí být takový, aby jeden balíček SIP obsahoval celý jeden objekt CDO (digitální originál) a všechna potřebná metadata. Postup vytváření balíčků AIP musí být takový, aby jeden balíček AIP vždy obsahoval celý jeden objekt CDO (jakožto odvozeninu digitálního originálu, vzniklou replikací) a všechna potřebná metadata. Archiv musí tato metadata doplnit o další interpretační informace a každá nová verze balíčku AIP musí obsahovat celý jeden objekt CDO (jakožto odvozeninu digitálního originálu, vzniklou transformací) a všechna potřebná metadata. Postup vytváření balíčku DIP musí být takový, aby balíček DIP byl vždy odvozován z aktuální (nejnovější) verze balíčku AIP a aby jeden balíček DIP vždy obsahoval celý jeden objekt CDO (jakožto odvozeninu digitálního originálu, vzniklou replikací, nebo transformací, v závislosti na požadavcích cílové komunity na prezentační formát).

#### **Stanovit rozdělení odpovědností v rámci organizace (P-0.5)**

Knihovna musí formálně ustanovit dva odlišné útvary, jeden pro digitalizaci a druhý pro archivace a zpřístupnění. Sloučení archivace a zpřístupnění do jednoho útvaru odpovídá potřebám požadavku „P-0.3“. Digitalizační útvary musí vykonávat provozní činnost a specializovanou činnost v oblasti nastavení snímacích zařízení a kontroly kvality digitalizace. Dále musí zahrnovat dostatečným počtem katalogizátorů, kteří pracují pro digitalizaci (kontrolují a doplňují katalogizační záznamy). Management archivu OAIS musí být managementem obou těchto útvarů, aby je mohl jednotně řídit, zejména z hlediska požadavků archivu na produkci (požadavek vytvářet digitální originál v archivačním formátu, který určí archiv apod.). Digitalizační projekt by měl být v souladu s celkovou koncepcí rozvoje fondů knihovny, která musí být vytvořena, podrobně zdokumentována a zpřístupňována čtenářům

## 8.2 Obecný postup pro transformace objektu CDO (sada „P-1“)

P-1.Obr	Stanovit a dodržovat pravidla transformace obrazových komponent	
	P-1.Obr.1	Stanovit typ transformace a její cíle
	P-1.Obr.2	Určit počet a sled dílčích operací jedné transformace
	P-1.Obr.3	Stanovit pracovní postup pro transformace, validace a metadatový záznam
	P-1.Obr.4	Provést testování pracovního postupu
	P-1.Obr.5	Vykonat transformaci a její validaci
	P-1.Obr.6	Zaznamenat transformace a její validaci do metadat
P-1.Str	Stanovit a dodržovat pravidla transformace strukturální komponenty	
P-1.Txt	Stanovit a dodržovat pravidla transformace textové komponenty	

Obecný postup pro transformace objektu CDO musí být dodržen v jakékoliv fázi životního cyklu. Je dále rozdělen na pravidla postupů pro jednotlivé digitální komponenty.

### **Stanovit pravidla transformace obrazové komponenty (P-1.Obr)**

Transformace obrazových komponent je nejsložitější a také nejrizikovější z hlediska udržení autenticity. Je proto podrobněji rozvedena do následujících dílčích požadavků.

#### **Stanovit typ transformace a její cíle (P-1.Obr.1)**

Tento požadavek znamená pojmenovat a popsat typ transformace v dokumentaci a stanovit její cíle formou specifikace digitální identity obrazové komponenty. Užití odlišného nástroje znamená odlišnou transformaci.

Určení typové i konkrétní digitální identity musí být v souladu s klíčovými vlastnostmi, které musejí být definovány před zahájením digitalizace (viz požadavek „P-2.1“ v oddíle 8.3). Typovou identitu musí tvořit nejméně formát (vyjádřený identifikátorem PUID a dále údajem o oficiálním názvu formátu a jeho verzi), formátový profil a vybrané obrazové vlastnosti (bitová hloubka, prostorové rozlišení, barevný model, barevný profil). Konkrétní digitální identitu musí tvořit minimálně pixelová velikost obrázku a název souboru (určený na základě zavedeného systému pojmenovávání pro generace obrazových dat).

#### **Určit počet a sled dílčích operací jedné transformace (P-1.Obr.2)**

Tento požadavek znamená zaprvé určit počet dílčích operací, které je vhodné nebo nutné zařadit do jedné transformace. Do jedné transformace je vhodné zařadit více operací podobného typu, pokud je vykonává tentýž transformační nástroj. Do jedné transformace je nutné zařadit operace, které je nutno vykonat všechny, aby došlo k dané transformaci, zejména z důvodu možností užitého nástroje (podrobnější vysvětlení s příklady bylo uvedeno v oddíle 7.4.2.2)

### **Stanovit pracovní postup pro transformace, validace a metadatový záznam (P-1.Obr.3)**

Tento požadavek znamená přiřadit ke každé operaci vhodný nástroj a sled, ve kterém budou provedeny v rámci jednoho nástroje. Validace znamená popsat všechny případné dílčí validace jednotlivých operací a jejich sled a zvolené validátory. Každá transformace (vyjma snímání) předpokládá, že k objektu CDO, který je předmětem transformace, existují technická metadata, ve kterých je zaznamenána digitální identita. V některých případech je nutná taková validace, která srovnává metadata předchozího objektu (např. konkrétní pixelovou velikost, která nesmí být změněna). V jiných případech validace předpokládá zadání hodnoty, která se má kontrolovat (např. identifikátor PUID pro nový formát). V dalších případech je nutné vytvořit mapovací tabulky mezi technickými údaji předchozího objektu a výsledného objektu, a jen na základě nich může být validace provedena.<sup>89</sup> V průběhu digitalizace není možné některé vlastnosti obrazových komponent zkontrolovat dříve než při finalizaci balíčku SIP (např. kontrola správnosti ořezů nebo vizuální kvality obrazu prováděna pracovníky digitalizačního útvaru při kontrole kvality – viz požadavek „P-2.7“ v oddílu 8.3). Zde musí být součástí informací o validaci také údaj o tom, kdy a kterým pracovníkem byla provedena tato manuální kontrola.

Dále musí stanoveny mechanismy metadatového záznamu pro každou transformaci. Jednak musí být určeny vhodné extraktory metadat a jednak musí být metadata plněna v souladu s daným metadatovým profilem. Stanovení událostí (ve smyslu standardu PREMIS) se musí odvíjet podle stanovených operací a validací.

### **Provést testování pracovního postupu (P-1.Obr.4)**

Tento požadavek je velmi důležitý z hlediska praktické realizovatelnosti. Musí být prokázána bezchybnost a účinnost navrženého postupu transformace (užitím validátorů), schopnost validátorů provést požadované validace a schopnost extraktorů plnit požadovaná metadata.<sup>90</sup> V případě nevhodnosti postupu nebo užitých nástrojů musejí být vybrány jiné a celý proces („P-1.Obr.1“ až P-1.Obr.1) se musí opakovat. Testování slouží k tomu, aby byl vytvořen nalezen optimální postup, u kterého lze konstatovat, že je funkční.

---

<sup>89</sup> Například pokud je typová digitální identita "matematicky bezeztrátová komprese" a má být zachována, ve metadatach zdrojového formátu TIFF může být jako údaj o kompresi uvedeno "LZW" a u výsledného formátu JP2 „JPEG 2000 Lossless“. Nejde o shodné údaje, ani o stejný kompresní algoritmus, ale v obou případech jde o matematicky bezeztrátovou kompresi.

<sup>90</sup> Může se například ukázat, že transformační nástroj není schopen vytvořit validní formát TIFF 6 nekomprimované verze, nebo že tuto schopnost má, ale zvolený validátor formát označí za nevalidní, ačkoliv validní je.

### **Vykonat transformaci a její validaci (P-1.Obr.5)**

Tento požadavek je možné splnit až poté, pokud po testování daný postup shledán za funkční. Znamená provedení transformace a její validace na všech souborech obrazové komponenty. Pokud proběhne je validace v pořádku, je nutno okamžitě vytvořit nový digitální otisk pro nové soubory.

### **Zaznamenat transformaci a její validaci do metadat (P-1.Obr.6)**

Tento požadavek znamená zaznamenat všechny události (operace v rámci transformace a validace) do metadat schématu PREMIS a rovněž do něj zaznamenat jejich pořadí a užití nástroje. Popis typu transformace (název zahrnující všechny operace) a její validace musí být také zaznamenán do dokumentace

### **Stanovit a dodržovat pravidla transformace strukturální komponenty (P-1.Str)**

Tento požadavek znamená jednak to, že soubory musejí být uloženy a pojmenovávány podle požadavků metadatového profilu. Názvy souborů musejí být jedinečné v celém digitalizačním projektu i v archivu. Z toho dále vyplývá nutnost na základě nových názvů souborů a jejich umístění aktualizovat strukturální komponentu (strukturální mapu v souboru ve formátu METS XML) a musí být validována správnost názvů a umístění souborů a jejich popisu ve strukturální komponentě. V daných případech může být nutná současná manuální kontrola (např. při kontrole kvality v digitalizaci (viz požadavek „P-2.7“ v oddíle 8.3).

### **Stanovit a dodržovat pravidla transformace textové komponenty (P-1.Txt)**

Tento požadavek znamená, že v digitalizaci je textová komponenta vytvořena jako výsledek jediné transformace a přiřazena k obrazové komponentě digitálního originálu. Požadavkem pro následné transformace v archivu a digitální knihovně je, aby byla zachována pixelová velikost souborů obrazové komponenty původního digitálního originálu, na kterou je textová komponenta vázána. V případě, že to není možné, je nutné v archivu vytvořit novou textovou komponentu (spustit proces OCR na obrazové komponentě v balíčku AIP znovu).

Důležitým údajem, který musí být uveden v metadatach textové komponenty, je údaj o obrazové komponentě, ze které byla textová komponenta vytvořena procesem OCR, a dalším důležitým údajem je údaj o procentuální úspěšnosti procesu OCR



### 8.3 Požadavky pro digitalizaci (sada „P-2“)

P-2.1	Stanovit a vyjádřit atributy identity digitalizátu knihy
P-2.2	Zvolit a zajistit vhodné nástroje pro digitalizaci
P-2.3	Stanovit postupy odvozování pro všechny digitální komponenty
P-2.4	Vytvořit digitalizační metadatový profil
P-2.5	Vytvořit dostatečný bibliografický záznam
P-2.6	Provést digitalizaci podle stanového postupu
P-2.7	Stanovit postupy pro kontroly kvality balíčku SIP
P-2.8	Stanovit postupy pro dodávání a opravy balíčků SIP

### **Stanovit a vyjádřit atributy identity digitalizátu knihy (P-2.1)**

Toto je základní požadavek pro celé řízení životního cyklu. Tyto atributy tvoří základ pro udržování a posuzování autenticity konkrétního digitalizátu knihy. Musejí být explicitně vyjádřeny v metadatech a dokumentaci, které musí být dostupné čtenářům.

P-2.1	Stanovit a vyjádřit atributy identity digitalizátu knihy			
	P-2.1.a	Vyjádření atributů tištěné předlohy digitalizátu knihy		
		P-2.1.a.1	Atributy exempláře knihy	
			P-2.1.a.1.1	Identifikátor exempláře knihy
			P-2.1.a.1.2	Název vlastníka exempláře knihy
			P-2.1.a.1.3	Údaj o stavu exempláře
			P-2.1.a.1.4	Údaj o akvizici
		P-2.1.a.2	Atributy vydání knihy	
			P-2.1.a.2.1	Perzistentní identifikátor vydání knihy
			P-2.1.a.2.2	Údaj o barevných vlastnostech knihy
			P-2.1.a.2.3	Minimální bibliografický záznam
		P-2.1.a.3	Údaj o časovém hledisku tištěné předlohy	
	P-2.1.b	Vyjádření atributů digitalizátu knihy		
		P-2.1.b.1	Perzistentní identifikátor digitalizátu knihy	
		P-2.1.b.2	Název vlastníka digitalizátu knihy	
		P-2.1.b.3	Popis klíčových vlastností digitalizátu knihy	
		P-2.1.b.4	Popis typové digitální identity digitálního originálu	
		P-2.1.b.5	Datum vzniku digitálního originálu	
		P-2.1.b.6	Název digitalizačního projektu	
	P-2.1.c	Vyjádření vztahu mezi digitálním originálem a tištěnou předlohou		
		P-2.1.c.1	Popis úrovně abstrakce digitalizace podle FRBR	
		P-2.1.c.2	Popis metody transformace	

### **Vyjádření atributů tištěné předlohy digitalizátu knihy (P-2.1.a)**

Tištěná předloha je součástí referenčního objektu autenticity digitalizátu knihy. Její atributy musejí být proto zaznamenány v přidružených metadatech nebo dokumentaci (jako informace popisující zdroj digitalizátu knihy).

Minimálními požadavky z hlediska atributů exempláře tištěné knihy jsou identifikátor exempláře knihy (P-2.1.a.1.1), název vlastníka exempláře knihy (P-2.1.a.1.2), údaj o stavu exempláře (P-2.1.a.1.3) a údaj o akvizici (P-2.1.a.1.4). Tyto údaje tvoří základ provenienčních informací o tištěné předloze jako o konkrétním fyzickém objektu. Údaj o stavu exempláře je rozhodný pro posuzování kvality obrazu digitálního originálu, údaj o akvizici je důležitý zejména z hlediska vysvětlení případů, kdy kniha (a její digitalizát) obsahuje údaje o jiné knihovně, než která je udávána jako vlastník exempláře.<sup>91</sup>

Minimálními požadavky z hlediska atributů vydání tištěné knihy jsou perzistentní identifikátor vydání knihy (P-2.1.a.2.1), údaj o barevných vlastnostech knihy (P-2.1.a.2.2) a minimální bibliografický záznam (P-2.1.a.2.3). Tyto informace jsou důležité z hlediska čtenářů, které primárně zajímá úroveň vydání knihy, nikoliv konkrétní exemplář. Perzistentní identifikátor je klíčový pro jednoznačné propojení digitalizátu knihy s jeho předlohou pro potřeby snadného vyhledávání čtenářů, celkovou koncepci rozvoje fondu knihovny (jako nástroj, jak bezpečně zjistit, zda již nějaká kniha nebyla zdigitalizována) a pro potřeby správy digitalizátů obecně. Minimální bibliografický záznam slouží jako informační doplněk k tomuto identifikátoru, který je důležitý pro potřeby čtenáře. Údaj o barevných vlastnostech musí zahrnovat popis jakýchkoliv barevných vlastností knihy (tj. včetně různé barvy textu) a slouží pro posuzování věrnosti digitalizačního převodu, spolu s údajem o časovém hledisku tištěné předlohy (P-2.1.a.3). Ačkoliv je výlučnou praxí snímat knihy ve stavu v době digitalizace, musí být tato informace zaznamenána. Ze všech atributů tištěné předlohy digitalizátu knihy jde o jediný údaj, který nelze zaznamenat do bibliografických metadat, ale měl by být zaznamenán do digitalizační dokumentace (vztahuje se na všechny knihy daného digitalizačního projektu).

### **Vyjádření atributů digitalizátu knihy (P-2.1.b)**

Perzistentní identifikátor digitalizátu knihy (P-2.1.b-1) je klíčový z hlediska dlouhodobé správy i identity digitalizátu v internetové síti. Musí splňovat obecné požadavky, které se očekávají od perzistentní identifikace digitálních dokumentů v internetovém prostředí

---

<sup>91</sup> Knihovna mohla knihu získat z vyřazených knih jiné knihovny, proto může obsahovat razítko jiné knihovny.

(viz oddíl 9.1.2) a dále musí umožnit odlišení všech možných digitalizátů téže předlohy (viz požadavky „PID-D“ v oddíle 9.2.1). Název vlastníka digitalizátu (P-2.1.b-2), údaj o datu vzniku digitálního originálu, tj. v balíčku SIP (P-2.1.b-5) a název digitalizačního projektu (P-2.1.b-6) jsou nezbytné provenienční informace o digitalizátu. Popis klíčových vlastností digitalizátu knihy (P-2.1.b-3) je nezbytný z hlediska dlouhodobého uchovávání digitalizátu (jako popis toho, co má být zachováno napříč transformacemi digitálního originálu) i zpřístupňování čtenářů (jako popis toho, co mohou čtenáři očekávat o digitalizátu prezentovaného v digitální knihovně). Za klíčové vlastnosti na úrovni intelektuální entity lze určit například tyto: plná barevnost, dobrá čitelnost textu, původní posloupnost, navigovatelnost, plnotextová prohledatelnost, kopírovatelnost textu. Popis typové digitální identity digitálního originálu (P-2.1.b-4) se vztahuje na digitální identitu objektu CDO v balíčku CDO, která má být vytvořena v rámci digitalizace, přičemž musí být v souladu v klíčových vlastnostech.

Tyto atributy musejí být zapsány do metadat v balíčku SIP (s možnou výjimkou klíčových vlastností). Popis typové digitální identity digitálního originálu a popis klíčových vlastností digitalizátu knihy musejí být zaznamenány v digitalizační dokumentaci před zahájením digitalizace. Klíčové vlastnosti pak musejí být také součástí dokumentace archivu.

### **Vyjádření vztahu mezi digitálním originálem a tištěnou předlohou (P-2.1.c)**

Tento požadavek zahrnuje popis zvolené úrovně abstrakce digitalizace podle FRBR (P-2.1.c.1), který je v současné praxi na úrovni exempláře a vydání s důrazem na vydání. Popis metody transformace (P-2.1.c.2) musí vycházet z údaje o časovém hledisku tištěné předlohy (P-2.1.a.3). Pokud je zvoleno snímání aktuálního stavu knihy, musí být metoda transformace taková, aby digitální originál co nejvěrněji zachovával současný vzhled knihy. Pokud je zvolena digitalizace do podoby předpokládaného stavu knihy v době vzniku, musí metoda transformace zahrnovat digitální restaurování a musí být uveden, jakým způsobem se zjistil tento předpokládaný stav. Všechny tyto informace musejí být zaznamenány v digitalizační dokumentaci před zahájením digitalizace. Do metadat je vyjádřit lze jen obtížně a nebylo by to ani účelné.

### **Zvolit a zajistit vhodné nástroje pro digitalizaci (P-2.2)**

Tento požadavek je specifikací požadavků na transformace objektu CDO (P-1). Specificky pro digitalizaci je nutno vybrat takové snímací zařízení, které dokáže vytvářet původní snímky v nekomprimovaném formátu TIFF nebo RAW a zapisovat EXIF metadata do souborů. V celém digitalizačním projektu musejí být užity stejné nástroje stejných verzí. Digitalizační projekt musí vytvořit vlastní komplexní validátor, který umožní komplexní validaci balíčku SIP (validaci všech obrazových komponent, metadat a názvů a umístění souborů).

### **Stanovit postupy odvozování pro všechny digitální komponenty (P-2.3)**

Tento požadavek znamená uzpůsobit obecné požadavky pro transformaci objektu CDO (P-1) pro potřeby digitalizace, zejména s ohledem na typovou identitu digitálního originálu a požadavky věrnosti. Základním pravidlem musí být, aby řetězec odvozování obrazové a textové měl stejný počátek (v původním snímku). Strukturální komponenta může být vytvořena až při vytváření balíčku SIP (kdy jsou již vytvořeny všechny ostatní soubory).

V případě, že projekt zvolí dva zásadně odlišné postupy transformací, musí stanovit dva celkové dílčí postupy. To platí pro případ, že je současně užito skenování a fotografování. V metadatech musí být zaznamenáno, jaký postup byl pro daný digitální originál užít, a to odkazem na postup uvedený v digitalizační dokumentaci. Při fotografování musí být vypočítáno předpokládané prostorové rozlišení a tento údaj vložen do obrazových souborů (z důvodu možnosti v budoucnu digitalizát vytisknout v původní velikosti předlohy).

### **Vytvořit digitalizační metadatový profil (P-2.4)**

Základem tohoto požadavku je užít pouze mezinárodní a široce užívané metadatové standardy. Popis specifických vlastností, které nelze užitím těchto standardů zaznamenat, by měl být uveden v digitalizační dokumentaci. Logika zápisu událostí musí vycházet z obecného postupu pro transformace (P-1). Metadatový profil musí také obsahovat řízené slovníky pro digitalizaci (zejména typologii události), systém pojmenovávání souborů a postup, jakými nástroji budou konkrétní elementy plněny (včetně určení verze těchto nástrojů). Tento postup musí být zachován pro celou digitalizaci. Reálně by měl tento profil vytvářet archiv (viz požadavek „P-3.1“ v oddíle 8.4)

### **Vytvořit dostatečný bibliografický záznam (P-2.5)**

Tento požadavek znamená provést před zahájením digitalizace knihu kontrolu a rozšíření bibliografického záznamu tak, aby obsahoval všechny požadované atributy exempláře knihy (P-2.1.a.1) a vydání knihy (P-2.1.a.2).

### **Provést digitalizaci podle stanového postupu (P-2.6)**

Tento požadavek znamená provedení digitalizace podle postupu, který musí být předem zdokumentován. Digitalizace musí před ostrým provozem provést testovací digitalizaci, v souladu s požadavkem („P-1.Obr.4“). Pokud se ukáže, že stanovený postup přináší chyby, musí být vytvořen nový postup.

### **Stanovit postupy pro kontroly kvality balíčku SIP (P-2.7)**

Tento požadavek zahrnuje nutnost disponovat komplexní validátorem balíčku SIP (viz požadavek „P-2.2“). Současně musí být provedena manuální kontrola struktury digitalizátu knihy (posloupnosti stran) a jeho vizuální kvality s tištěnou předlohou v ruce.

### **Stanovit postupy pro dodávání a opravy balíčků SIP (P-2.8)**

Tyto postupy musí být založeny na principu, že v případě větších chyb balíčku SIP musí být provedena opětovná digitalizace. Za větší chyby je nutno považovat nekvalitní obraz a neúplnost digitalizátu.

## 8.4 Požadavky pro archivaci (sada „P-3)

P-3.1	Stanovovat požadavky na balíčky SIP
P-3.2	Vybírat nebo vyvíjet vhodné nástroje pro archivaci
P-3.3	Vytvořit a udržovat archivační metadatový profil
P-3.4	Stanovovat postupy pro vytváření balíčků AIP
P-3.5	Shromáždit a uchovávat kompletní digitalizační dokumentaci
P-3.6	Stanovit postupy plánování uchovávání
P-3.7	Provádět archivační opatření
P-3.8	Vytvářet a udržovat dokumentaci archivu
P-3.9	Stanovovat postupy pro opravy

### **Stanovovat požadavky na balíčky SIP (P-3.1)**

Tento požadavek znamená, že archiv by měl mít rozhodující slovo při stanovování podoby digitálního originálu. Zejména to znamená volbu archivačního formátu a vytvoření digitalizačního metadatového profilu.

### **Vybírat nebo vyvíjet vhodné nástroje pro archivaci (P-3.1)**

Tento požadavek zahrnuje zejména výběr metadatových extraktorů a nástrojů pro formátovou identifikaci a validaci. Tyto nástroje musejí být v průběhu času nahrazovány novějšími verzemi nebo jinými nástroji, v souvislosti s vývojem technologického prostředí. V současnosti musí využívat nástroje JHOVE, jpylyzer a DROID a doplnkově další nástroje. Všechny nástroje musejí být integrovány do celkového softwarového řešení archivu. V případě, že je známo, že doporučované nástroje přinášejí odlišné výsledky a neexistuje jednoznačné řešení, je nutno, aby tato skutečnost byla zdokumentována a vyřešena v budoucnosti. Archiv také musí v budoucnosti zajistit vhodné transformační nástroje pro formátovou konverzi.

### **Vytvořit a udržovat archivační metadatový profil (P-3.3)**

Archivační metadatový profil se na rozdíl od digitalizačního bude v průběhu času měnit, v souvislosti s novými verzemi metadatových standardů nebo novými standardy. Softwarový systém musí být navržen tak, aby jednak umožňoval aplikaci tohoto

metadatového profilu, jednak jeho úpravy v budoucnosti. Archivační metadatový profil se tedy musí vyvíjet, zatímco digitalizační musí zůstat stejný po celou digitalizaci.

#### **Stanovovat postupy pro vytváření balíčků AIP (P-3.4)**

Tyto postupy musejí zahrnovat kompletní validaci objektu CDO v balíčku SIP (včetně kontroly neporušenosti) v souladu s požadavky na transformace (P-1). Dále musejí zahrnovat postupy pro shromažďování interpretačních informací k objektu CDO v balíčku SIP (bez ohledu na to, že již byly částečně shromážděny při digitalizaci) a v průběhu času také k objektu CDO v balíčku AIP. Metadata vytvořená v balíčku SIP musejí být v balíčku AIP uložena odděleně od metadat vytvářených archivem (tj. v jiných souborech) a musí být jasně označeno, jaká metadata byla vytvořena v digitalizaci a jaká při archivaci.

Součástí tohoto požadavku je nutnost zachovat první verzi balíčku AIP (po vytvoření nové verze balíčku AIP) jako součást dokumentace o autenticitě.

#### **Shromáždit a uchovávat kompletní digitalizační dokumentaci (P-3.5)**

Tento požadavek znamená, že archiv musí v průběhu digitalizace shromáždit a následně uchovávat kompletní digitalizační dokumentaci, případně nechat dopsat chybějící části. Tato dokumentace musí zahrnovat digitalizační aplikační profil, vyjádření všech atributů identity intelektuální entity, popis postupů odvozování a popis všech užitých nástrojů.

Archiv rovněž musí shromažďovat všechny užití metadatové standardy všech verzí, formátové specifikace a užití specializované nástroje (extraktory metadat, validátory formátů apod.) a jejich dokumentaci a uchovávat je spolu s digitalizáty.

#### **Stanovit postupy plánování uchovávání (P-3.6)**

Tento požadavek se týká činnosti archivu v dlouhodobém horizontu. Archiv musí alespoň jednou ročně provést posudky uložených formátů (z hlediska zastarávání), nových verzí metadatových standardů (z hlediska možného užití v archivačním metadatovém profilu) a nových verzí specializovaných nástrojů nebo nových nástrojů. V případě, že bude vydán metadatový standard pro záznam klíčových vlastností, musí jej začlenit do archivačního metadatového profilu a metadata všech balíčků AIP obohatit o tyto údaje.

V budoucnosti, až dojde k zastarání archivačního formátu, musí archiv po pečlivém studiu odborné literatury a osvědčené praxe významných institucí zvolit nový archivační formát, vhodné transformační nástroje a tuto volbu důkladně zdokumentovat.



### **Provádět archivační opatření (P-3.7)**

Hlavním předpokládaným archivačním opatřením je formátová konverze, která musí být v budoucnosti provedena v souladu s postupy plánování uchovávání (P-3.6). Požadavky na tyto konverze musejí vycházet z obecných postupů pro transformace objektu CDO (P-1), s tím rozdílem, že musí být proveden mnohem důkladnější průzkum vhodných formátů a nástrojů (včetně konzultace s jinými organizacemi) a celý proces musí být důkladně zdokumentován, včetně odůvodnění tohoto kroku. Po skončení formátové konverze musí být vytvořen souhrnný protokol, popisující celý postup, dotčenou množinu objektů CDO a následné validace.

### **Vytvářet a udržovat dokumentaci archivu (P-3.8)**

Dokumentace archivu musí být průběžně doplňována a dokumentovat postupy uvedené ve všech ostatních požadavcích sady „P-3“.

### **Stanovovat postupy pro opravy (P-3.9)**

Tento požadavek se musí řídit principem, že každá oprava, která mění objekt CDO, musí být provedena tím způsobem, že po jejím vykonání bude vytvořen nový balíček AIP (nový balíček AIP) a původní balíček bude buď okamžitě smazán, nebo deaktivován (tj. popsán tak, aby při budoucí replikaci na nový nosič byl smazán). Pro udržení vazby mezi původním balíčkem AIP a novým balíčkem AIP musí být užit perzistentní identifikátor (blíže viz požadavky „PID-D“ v oddíle 9.2.2).

## 8.5 Požadavky pro digitální knihovnu (sada „P-4)

P-4.1	Stanovovat vhodné prezentační formáty
P-4.2	Vyvíjet softwarové řešení digitální knihovny
P-4.3	Prezentovat digitalizát knihy spolu s atributy jejich identity
P-4.4	Nabízet doplňkové alternativní prezentační varianty ke stažení
P-4.5	Umožnit cílové komunitě stažení celého balíčku AIP
P-4.6	Spolupracovat se systémem perzistentní identifikace

Tyto požadavky se týkají fáze zpřístupňování a vyžadují splnění požadavku řídit digitální knihovnu jako součást archivu OAIS (P-0.3)

### **Stanovovat vhodné prezentační formáty (P-4.1)**

Tento požadavek znamená, že digitální knihovna musí volit takové prezentační formáty, které se dají zobrazit v internetovém prohlížeči bez nutnosti instalace pluginu. Součástí tohoto požadavku může být nutnost vytvořit prezentační meziformát, ze kterého se bude za chodu generovat prezentační formát (v současnosti je optimální volbou užití meziformátu JP2 pro generování prezentačního formátu JPEG). Součástí tohoto požadavku je nutnost vytvářet prezentační formáty (meziformáty) v digitální knihovně.

### **Vyvíjet softwarové řešení digitální knihovny (P-4.2)**

Tento požadavek znamená, že softwarové řešení digitální knihovny musí být vyvíjeno zejména v závislosti na vývoji internetových prohlížečů, požadavcích čtenářů a změnách zpřístupňovacího softwaru. Digitální knihovna musí vybírat a testovat nástroje, které umožňují reprodukovat informační obsah bezchybně a v souladu s klíčovými vlastnostmi.

### **Prezentovat digitalizát knihy spolu s atributy jejich identity (P-4.3)**

Tento požadavek tvoří jádro požadavků autenticity z hlediska čtenáře. Digitalizát knihy musí být zobrazen spolu s metadaty a odkazy na dokumentaci, které popisují všechny nezbytné atributy jeho identity (stanovené v požadavku „P-2.1). Tyto atributy musejí být vyjádřeny v úplnosti a bezchybně, aby čtenář nebyl uveden v omyl.

### **Nabízet doplňkové alternativní prezentační varianty ke stažení (P-4.4)**

Digitální knihovna musí také nabízet digitalizáty knih ke stažení v podobě jednoho souboru (např. formátu PDF). V tomto souboru musí být v úvodu uvedeny všechny

požadované atributy identity tohoto digitalizátu (s odkazy na dokumentaci pro atributy, které nelze vyjádřit v metadatech). Digitální knihovna dále musí uvádět informace o vhodných nástrojích k zobrazení tohoto souboru, informaci o digitálním otisku (pro možnost ověření neporušenosti) a případné informace o tom, že tato prezentační varianta je v nižší kvalitě, než jsou digitalizované knihy zobrazované v internetovém prohlížeči.

#### **Umožnit cílové komunitě stažení celého balíčku AIP (P-4.5)**

Digitální knihovna musí umožnit čtenářům (alespoň za určitých podmínek nebo alespoň vybranému okruhu čtenářů, např. vědců) stažení celého balíčku AIP, včetně všech metadat popisujících jeho provenienci.

#### **Spolupracovat se systémem perzistentní identifikace (P-4.6)**

Tento posední požadavek souvisí s potřebou perzistentní identifikace nezávislé na URL adrese (blíže viz oddíl 9.1.2.5). Digitální knihovna musí pro každý digitalizát vytvořit úvodní stránku, ze které budou přístupné všechny jeho části a která bude obsahovat metadata o attributech jeho identity. Dále musí zajistit, aby tuto URL adresu mohl pravidelně získávat identifikační systém. Tento požadavek tak souvisí s požadavky perzistentní identifikace digitalizátů knih (sada „PID“), které jsou uvedeny v oddíle 9.2.2.

## **9 Návrh pravidel perzistentní identifikace digitalizátů knih**

Tato kapitola je členěna do dvou částí. První část (podkapitola 9.1) představuje analýzu základních požadavků na perzistentní identifikátory a jejich systémy. Informace zde uvedené jsou výsledkem zobecnění, které jsem provedl na základě analýzy fungování úspěšně zavedených identifikačních systémů přidělujících perzistentní identifikátory v oblasti kulturního dědictví. Perzistentní identifikace je těmito systémy již řešena delší dobu (původně pro oblast fyzických, následně i digitálních objektů) a její základy jsou obsaženy v osvědčených postupech těchto systémů. Je tedy nutné z těchto postupů, po určité revizi, vycházet a na základě nich specifikovat požadavky pro perzistentní identifikaci digitalizátů knih. Výsledky analýzy prezentované v podkapitole 9.1 již byly částečně publikovány [(4), (5) a (3)].

Druhá část (podkapitola 9.2) obsahuje specifické požadavky pro perzistentní identifikaci digitalizátů knih, které jsou součástí širších požadavků pro udržení a posouzení jejich autenticity, a které jsou výsledkem plnění jednoho z cílů této disertační práce.

### **9.1 Základní požadavky pro perzistentní identifikaci digitálních objektů**

#### **9.1.1 Úvod**

Identifikátor v obecném významu je znakovým prostředkem, který jednoznačně označuje (tj. identifikuje) určitý objekt (referent) v určitém kontextu. Kontextem se rozumí především oblast užití identifikátoru. Identifikátory mohou fungovat pouze v rámci určitého systému: identifikátor je prvkem určitého systému (identifikačního systému). Pro kontext této práce jsou relevantní systémy pro perzistentní identifikaci, jejichž cílem je identifikace publikací a dokumentů pro potřeby lidských uživatelů. Za nejvýznamnější systémy v této oblasti lze jednoznačně považovat systémy ISBN (99), ISSN (100), Handle (101), DOI (102) a URN:NBN (103).

Systém ISBN vznikl na přelomu 60. a 70. let 20. století. Jeho primárním cílem bylo zjednodušit komunikaci mezi vydavateli, distributory, knihkupci a čtenáři (104, s. 6-7). Identifikátor ISBN se také stal osvědčeným perzistentním identifikátorem pro potřeby knihoven. Namísto mnoha popisných údajů stačilo uvést deset čísel, čímž se odstranila mnohoznačnost jiných identifikačních údajů (viz problém stejných jmen autorů nebo stejných názvů knih apod.). Systém ISBN později rozšířil délku identifikátoru (na třináctimístné číslo) a umožnil také identifikaci elektronických publikací nabízených vydavateli. Za jednu z výhod

systému ISBN lze považovat skutečnost, že identifikátor ISBN je vždy součástí referentu (je vytištěn v knize), čímž je zpevněna vazba mezi identifikátorem a jeho referentem. Další výhodou je systém ohlášených knih (založený právě na identifikátoru ISBN, který je již před vydáním spojen s knihou), který umožňuje knihovnám připravit se na chystaná vydání a zjednodušit proces zpracování nové akvizice.

Krátce po vzniku systému ISBN byl vytvořen systém ISSN pro identifikaci tištěných periodických publikací, a následně také elektronických periodických publikací. Systémy Handle, DOI a URN:NBN vznikly jako nové systémy identifikace, zaměřené specificky na identifikaci digitálních publikací a dokumentů v internetové síti. Tyto systémy se v současné praxi většinou užívají pro ty publikace a dokumenty, kterým nelze přidělit identifikátor ISBN.

Všechny uvedené systémy lze popsat jako systémy, které jsou tvořeny následujícími prvky: identifikátory, soubor pravidel, účastníci, uživatelé, informace o identifikovaných objektech a organizační a technické zajištění systému. Pravidla systému mj. definují strukturu identifikátoru, podmínky a možnosti užití identifikátoru (z nichž klíčové je stanovení toho, jakým objektům lze identifikátor přidělit a jakým nelze) a formu kooperace mezi účastníky systému. Účastníky systému můžeme podle vykonávaných rolí rozdělit do dvou skupin: na registrační agentury a registrátory. Registrační agentury lze rozdělit na centrální a lokální. Centrální agentury rozhodují o pravidlech systému, lokální agentury zpravidla přidělují identifikátory v dané zemi nebo pro daný okruh uživatelů. Například u systému ISBN jde o dělení na mezinárodní agenturu ISBN a národní agentury ISBN, které přidělují identifikátory ISBN vydavatelům (národní agentury jsou většinou národní knihovny). Totéž členění platí pro systém ISSN určený k identifikaci periodik. V případě systému DOI jde o dělení na Mezinárodní nadaci pro DOI (International DOI Foundation) jako centrální agenturu a několik podřízených agentur, které působí nadnárodně a přidělují identifikátory DOI obvykle jednotlivým vydavatelům. V případě systému URN:NBN centrální agentura existuje virtuálně, a to smyslu, že pravidla systému vymezuje standard URN:NBN, který také určuje, že lokálními agenturami mohou být pouze národní knihovny, přičemž systém neřídí žádná centrální organizace (103). V případě systému Handle existuje pouze centrální agentura, kterou představuje americká Korporace pro národní výzkumné iniciativy (Corporation for National Research Initiatives), která přiděluje jmenné podprostory identifikátorů Handle registrátorům tohoto systému (101). Registrátoři jsou organizace, které nechávají přidělovat identifikátory zpravidla publikacím, které vydávají, nebo dokumentům, které spravují. Registrátory systémů ISBN a ISSN jsou vydavatelé neperiodických, resp.

periodických publikací, registrátory systému DOI většinou komerční vydavatelé odborných periodik (identifikátorem DOI se většinou označují články těchto periodik), registrátory systému Handle různé typy organizací (mj. vysoké školy) a registrátory systému URN:NBN většinou vysoké školy a výzkumné organizace (3).

Uživateli identifikačního systému jsou myšleni koncoví uživatelé, zejména čtenáři. Koncoví uživatelé nejsou účastníky identifikačního systému identifikace, ale využívají identifikátory pro usnadnění práce s informačními zdroji (např. pro vyhledávání v katalogu knihovny, objednávky u vydavatelů apod.). Koncové uživatele lze považovat za hlavní cílovou skupinu identifikačních systémů. Například identifikátory ISBN sice využívají pro komunikaci mezi sebou vydavatelé, distributoři, knihkupci i knihovny, ale kontext užití těchto identifikátorů je prakticky nemyslitelný bez čtenářů, kteří jsou hlavní cílovou komunitou těchto subjektů.

Organizační a technické zajištění je nezbytnou součástí každého komplexního identifikačního systému (blíže se mu budeme věnovat v oddíle 9.1.3).

V následujících oddílech této podkapitoly se věnujeme specifickým otázkám perzistentní identifikace digitálních objektů, a to ze dvou hledisek: z hlediska identifikátoru (oddíl 9.1.2) a z hlediska systému (oddíl 9.1.3) Perzistentní identifikátory digitálních objektů budeme pro zjednodušení nazývat digitální identifikátory.

### **9.1.2 Obecné požadavky na perzistentní digitální identifikátory**

Digitální identifikátor může plnit řadu funkcí. V následujícím přehledu uvedeme takové, které jsou klíčové z hlediska dlouhodobého uchovávání i zpřístupňování digitálních objektů čtenářům. Požadavky z kategorie jednoznačnosti platí digitální identifikátory i identifikátory fyzických objektů (a také některé dílčí požadavky kategorií facilitace, perzistence a globálnosti). Požadavek trvalého přesměrovávání je specifickým požadavkem pro digitální identifikátory.

#### **9.1.2.1 Jednoznačnost identifikace**

Základní funkce identifikátoru, tj. jednoznačně identifikovat nějaký objekt, je nejčastěji zmiňovanou funkcí identifikátoru, neboť je jeho základním definičním prvkem.

Jednoznačná identifikace znamená odstranění mnohoznačnosti typické pro běžný jazyk. Z tohoto důvodu jsou identifikátory užívané ve sledovaných komplexních identifikačních systémech většinou takové řetězce znakové sady, které jsou uměle vytvořené

(tj. pro koncového uživatele bezvýznamové, například sled čísel identifikátoru ISBN). V praxi to obvykle znamená, že v řadě procesů je identifikátor dostačujícím prostředkem pro jednoznačnou identifikaci referentu. Například pro vyhledávání stačí, aby čtenář zadal pouze identifikátor, aniž by musel vyplňovat různé další selekční údaje.

Jednoznačnost znamená vždy jednoznačnost v rámci daného systému. Registrátoři i uživatelé musejí znát kontext užití identifikátoru. Pro registrátory to znamená povinnost řídit se pravidly daného systému, pro čtenáře rozumět tomu, že například identifikátory ISBN se přidělují knihám a tomu přizpůsobit svoje očekávání. Jednoznačnost však neznámá, že tentýž objekt nemůže mít více identifikátorů podobného typu. Taková skutečnost není na škodu, naopak. Například pravidla systémů ISBN a ISSN umožňují některých typům publikací přidělit identifikátor ISBN i ISSN. Existence více systémů identifikace může teoreticky zvýšit pravděpodobnost jednoznačné identifikace. Témuž objektu však nelze přidělit dva různé identifikátory stejného systému. To souvisí s otázkou perzistence, o které je pojednáno v oddíle 9.1.2.3.

### **9.1.2.2 Facilitace užití identifikátoru**

Funkce jednoznačné identifikace je v praxi úzce spjata s druhou funkcí identifikátoru, jíž je facilitace operací s identifikovaným objektem, tj. usnadnění práce s identifikovaným objektem. Tyto operace musejí zahrnovat v kontextu řízení celého životního cyklu otázky zpracování, dlouhodobé správy a zpřístupňování (vyhledávání, citování apod.). Identifikátor pak zde může sloužit jako zástupný prvek za velké množství jiných údajů (např. při komunikaci bibliografických záznamů nebo při vyhledávání identifikátor nahrazuje různé popisné údaje, které by jinak musely být užity při výměně záznamů nebo při vyhledávání). Z toho důvodu všechny komplexní identifikační systémy stanovují takovou podobu identifikátoru, která se řídí jednoduchým ekonomickým principem: struktura samotného identifikátoru (tedy znakového řetězce) musí být na jedné straně co nejkratší (aby se lidem s identifikátory co nejlépe pracovalo) a současně schopna pojmut předpokládaný rozsah identifikovaných referentů v daném kontextu.

Nalezení rovnováhy mezi oběma maximami nemusí být při projektování identifikačních systémů vždy jednoduchou záležitostí. Přejít systému ISBN z desetimístného znakového řetězce na třináctimístný v roce 2007 ukazuje, jak obtížné je tento ekonomický princip navrhnout s ohledem na predikci budoucího vývoje. V tomto případě nejde jen o otázku nárůstu celkové produkce knižních titulů, ale též důsledky technologického vývoje (publikace je možno vydat v různých formátech a publikace

v odlišném formátu musí mít podle pravidel ISBN přidělen jiný identifikátor, neboť je systém ISBN považuje za jiný referent).

Dalším aspektem facilitace je užití pouze omezené sady znaků. Například v systému URN:NBN je možno užít pouze alfanumerické znaky a dále spojovník a dvojtečku, které slouží jako významotvorné syntaktické znaky. Podobná omezení nalezneme i v dalších systémech. Pro koncové uživatele může být problém vyhledávat podle identifikátoru obsahujícího specifické nealfanumerické znaky. Do oblasti facilitace lze také zařadit možnost automatického ověření validnosti struktury identifikátoru, ať užitím kontrolní čísla (jako u identifikátoru ISBN) nebo kontrolou předepsané syntaxe.

### **9.1.2.3 Perzistence identifikátoru**

Požadavek perzistence identifikátoru znamená, aby identifikátor jednoznačně identifikoval daný objekt v dlouhodobém horizontu. To je tradičním požadavkem knihoven, ale i celé řady jiných organizací a kontextů užití. To znamená, že jeden a tentýž identifikátor, který byl v určitém okamžiku přidělen určitému (podle pravidel identifikačního systému stanovenému) objektu, již nikdy nesmí být přidělen znovu žádnému jinému objektu.

Na požadavku perzistence je od samého začátku založen i systém ISBN, který je určen pro vydavatele. Pravidla systému tento požadavek vyjadřují zcela jasně: „Jakmile bylo jedno ISBN přiděleno monografické publikaci, nesmí být již nikdy použito k tomu, aby identifikovalo jinou monografickou publikaci, a to ani tehdy, pokud bylo toto původní ISBN přiděleno chybně.“ (104, s. 13). Tento příklad také ukazuje, že perzistence identifikace není vyžadována pouze v kontextu dlouhodobého uchovávání digitálního dědictví, ale i v komerčním sektoru.

Perzistenci nelze považovat za čistě technický parametr. Do značné míry jde totiž o otázku organizační, tedy stanovení vhodných pravidel a jejich dodržování všemi účastníky systému. Důležité je rovněž to, aby pravidla určila spolehlivé mechanismy, které zajistí, aby nedošlo k opětovnému přidělení již přiděleného identifikátoru jinému objektu. Zejména v kontextu digitálních identifikátorů je rozhodným požadavkem, aby registrační agentury udržovaly registr již přidělených identifikátorů, tedy vhodně zvolené databázové řešení, včetně mechanismů zálohování. Na tyto registry identifikátorů se tak de facto vztahují stejné požadavky jako na dlouhodobé uchovávání jimi označovaných objektů. V praxi jsou však zaznamenány případy, kdy identifikační systém není vhodně navržen pro splňování těchto požadavků. Například ve finském systému URN:NBN, který provozuje finská národní knihovna, si může registrátor nechat přidělit identifikátor URN:NBN online generátorem



provozovaným touto knihovnou jakožto registrační agenturou pro finské organizace. Tento generátor funguje jako jeden postupně inkrementovaný čítač, který si ukládá mechanismus přidělování. V původní verzi generátor potřeboval ukládat tento mechanismus pro celý rok, v současné verzi pouze pro jeden den. V obou případech však v případě výpadku hrozí, že nebude vygenerován jedinečný identifikátor (3, s. 23-25). Takto přidělené identifikátory jsou navíc dodány do registru finského identifikačního systému až později, do té doby o nich systém neví.<sup>92</sup> Rovněž není optimálním řešením, aby se generování identifikátorů ponechávalo na odpovědnosti registrátorů, jako je tomu například v nizozemském systému URN:NBN (3, s. 33).

V oblasti systémů perzistentní identifikace je tedy validnost identifikátoru posuzována ve dvou stupních: a) validnost struktury (skutečnost, že identifikátor odpovídá předepsané podobě a syntaxi), b) validnost perzistence (skutečnost, že nově přidělený identifikátor již nebyl v minulosti přidělen jinému objektu, a také že tento identifikátor je trvale uchováván v identifikačním systému).

Perzistence identifikátoru je také neodlučně spjata s konceptem perzistence objektu, který je označen perzistentním identifikátorem. Objekt, u kterého se nepředpokládá, že bude užíván (zpřístupňován, archivován, distribuován apod.) v dlouhodobém horizontu, nemá smysl označovat perzistentním identifikátorem. Dokumenty pomíjivé hodnoty není účelné identifikovat perzistentními identifikátory uvedených komplexních systémů.

#### **9.1.2.4 Globálnost identifikátoru**

V souvislosti s globalizací světa se objevil požadavek na další funkci identifikátoru – jeho mezinárodní jedinečnost. Globální systém identifikace vyžaduje takovou spolupráci při jeho budování a udržování, která přesahuje národní kontext jednoho státu. V případě digitálních objektů, které jsou určeny ke zpřístupňování v internetové síti, je navíc vysoce nepraktické, aby byl budován takový systém identifikace, jehož cílem by nebyla globálnost identifikátoru. Identifikační systémy s globálním dosahem se obvykle stávají mezinárodním standardem nebo normou, což je případ všech zkoumaných komplexních systémů. Prvním z nich v tomto ohledu byl systém ISBN, následovaný systémem ISSN.

---

<sup>92</sup> Registrátor dodá takto vygenerované identifikátory URN:NBN do registru systému až v rámci dodávky URL adres, na kterých se identifikované objekty nacházejí.

### 9.1.2.5 Trvalé přesměrovávání

Masová komputelizace a internetizace společnosti, která začala v 90. letech 20. století, zásadně změnila způsoby zpřístupňování informačních zdrojů. Internet rozšiřuje možnosti šíření knih nebo vědeckých článků, přičemž některé tradiční formy distribuce zanikají. Paralelně s tímto vývojem se logicky vynořila poptávka po nové funkci perzistentního identifikátoru: trvalé zprostředkovávání okamžitého přístupu k informačnímu zdroji. URL adresa, klasický internetový identifikátor (a pravděpodobně nejužívanější identifikátor v oblasti informačních zdrojů vůbec), se totiž ukázala být nestabilní nejen v dlouhodobém, ale i střednědobém a často i krátkodobém horizontu. Podle jednoho z novějších odhadů je průměrná životnost URL adresy sto dní (105, s. 6). Ačkoliv jde o paušální odhad a lze předpokládat, že URL adresy například digitálních knihoven mají delší trvání, URL adresa není vhodným perzistentním identifikátorem objektu, jak ostatně ukazuje běžná citační praxe (například několik zdrojů citovaných touto disertační prací již není dostupných).

URL adresa je pouze lokátorem objektu (dočasným identifikátorem umístění objektu v internetové síti). Systém Handle, a následně systémy DOI a URN:NBN, zavedly koncept perzistentních identifikátorů digitálních objektů nezávislých na URL adrese. Ten je založen na principu trvalého propojování perzistentního digitálního identifikátoru s aktuální URL adresou a přesměrováváním uživatele na tuto adresu, které zajišťuje resolver, základní služba těchto systémů. Uživatel tak potřebuje znát pouze perzistentní identifikátor a URL adresu, na které je resolver dostupný. Takto koncipovaný identifikační systém nemá analogii ve světě tradičních dokumentů (zpřístupňování tištěných knih v knihovnách atd.), kde užitím identifikátoru (např. v katalogu) lze okamžitě získat pouze metadata.

### 9.1.3 Obecné požadavky na identifikační systémy

#### 9.1.3.1 Technická infrastruktura

Technickou infrastrukturu uvedených identifikačních systémů pro digitální identifikátory (DOI, Handle, URN:NBN) můžeme konceptuálně rozdělit na tři části: centrální subsystém (provozovaný registrační agenturou), subsystémy registrátorů a mechanismy jejich komunikace. Jádrem technické infrastruktury pak lze rozdělit na čtyři základní komponenty (funkční celky): resolver, registr, generátor a evidenční služba.<sup>93</sup> Tyto komponenty jsou

---

<sup>93</sup> Původně vypracováno autorem této práce pro článek o systémech URN:NBN v Evropě (3). Tuto dekompozici lze však užít i pro systémy Handle a DOI.

konceptuální, v praxi je může vykonávat jeden nástroj. Resolver je komponenta, která zajišťuje dereferenci. Dereference je proces, při kterém je identifikátor vstupem (požadavkem) do síťové služby, která jako odpověď poskytne specifický výstup (106, s. 2). Výstupem dereference je primárně přesměrovávání na aktuální URL adresu objektu. Doplňkově však resolvers jako výstup dereference mohou poskytovat metadata o referentu, nebo o samotném identifikátoru (např. kdy byl přidělen). Resolver je vždy součástí centrálního subsystému. Registr musí mít k dispozici informace potřebné k dereferenci identifikátorů (minimálně databázi vazeb perzistentních identifikátorů a URL adres) a musí být spojen s resolverem. Generátor je nástroj pro generování perzistentních identifikátorů, který musí zajistit jejich jedinečnost. Generátor může být součástí centrálního subsystému, nebo mohou registrátoři využívat vlastní. Pokud registrátoři využívají vlastní generátory, musejí mít vlastní registry identifikátorů nebo jiné mechanismy pro zajištění jedinečnosti identifikátoru. Evidenční služba zajišťuje v první řadě získávání aktuálních URL adres, případně dalších údajů (např. samotných perzistentních identifikátorů, pokud si je generuje registrátor sám vlastním nástrojem, a teprve pak je dodává do centrálního subsystému spolu s URL adresou). Může být realizována například prostřednictvím protokolu OAI-PMH, který sklízí potřebné údaje (v první řadě perzistentní identifikátory a aktuální URL adresy) z digitální knihovny registrátora.

Tyto komponenty musejí být jednotně řízeny, buď centrálním subsystémem, nebo pravidly, která účastníci dodržují při užívání vlastních generátorů. Teoreticky nejvyšší spolehlivost bude mít identifikační systém, ve kterém jsou všechny tyto komponenty součástí centrálního subsystému a existuje jediný resolver. Centrální subsystém v tomto případě sám generuje identifikátory pro registrátory, tyto identifikátory jsou okamžitě poté uloženy do registru a následně evidenční služba získá od registrátora aktuální URL adresu a nakonec resolver nabídne čtenáři dereferenci. Evidenční služba by poté měla v odpovídajících intervalech aktualizovat URL adresy za nezbytné součinnosti na straně registrátora (optimálně by měl registrátor upozornit registrační agenturu na změnu URL adresy sám).

### **9.1.3.2 Organizační zajištění**

Jak je zřejmé z výše uvedeného, kvalitní identifikační systém je především založen na kvalitním organizačním zajištění. Musejí být zvolena vhodná pravidla jak pro užívání perzistentních identifikátorů, tak pro správu technické infrastruktury, která předpokládají aktivní zapojení všech účastníků, nejen registračních agentur. Softwarové řešení

je bezpochyby nezbytnou technickou páteří celého systému, ale nelze se spoléhat se na to, že software může nahradit lidské pracovníky na straně účastníků systému.

Celý identifikační systém nemůže provozovat jediná organizace. Registrační agentury potřebují součinnost na straně organizací registrátorů. Důvěryhodnost celého identifikačního systému se může odvíjet od toho, do jaké míry jeho pravidla dodržují jednotliví účastníci, tedy jak pečlivě dodávají aktuální URL adresy do centrálního subsystému, jakým způsobem uvádějí perzistentní identifikátor spolu s identifikovaným dokumentem, jak zajišťují dlouhodobé uchovávání identifikovaného objektu apod.

Organizační zajištění je nejnáročnější na straně registrační agentury, zejména z hlediska provozu centrálního subsystému. Systémy Handle a DOI jsou zpoplatněny. Systém URN:NBN, který provozují národní knihovny, zpoplatněn není. Jeden z důvodů, proč řada evropských národních knihoven zavedla tento systém a nevyužila služeb starších systémů Handle a DOI, je právě finanční hledisko.

### **9.1.3.3 Pravidla systému**

Identifikační systémy se liší mírou komplexnosti požadovaných pravidel. Základním pravidlem je dodávat do systému aktuální URL adresy. Relativně nejvolnější jsou pravidla systému Handle: identifikátor lze užít pro jakýkoliv referent (101). Jednotlivé národní systémy URN:NBN mají obvykle přísnější pravidla stanovující, jakým typům objektů lze přidělit identifikátory, do jaké míry granularity apod. (3). Nejkomplexnější pravidla má systém DOI, která zahrnují vlastní datový model a vyžadují dodání celé řady metadat vztahujících se k referentu (102). Systém ISBN stanovuje nejprísnejší model referentu. Tento model v pravidlech systému není popsán systematicky (104). Z dostupného popisu však lze tento model formulovat následovně. Typem dokumentu musí být monografická publikace. Intelektuálním obsahem musí být jednotlivé dílo (tj. jakýkoliv intelektuální výtvar, literární, vědecký, naučný apod.), které nemá pomíjivou hodnotu, je jako samostatný celek vydáno a distribuováno širší komunitě uživatelů (zpravidla čtenářů) na komerční nebo nekomerční bázi a je primárně textové povahy. Z hlediska formy pak platí, že každá odlišná forma vydání (vázané vydání, brožované vydání, formát PDF apod.) zakládá nový objekt identifikace, kterému musí být přidělen jiný identifikátor, i když se jedná o tentýž intelektuální obsah. Z hlediska vydavatele pak platí, že jiný vydavatel nebo jiné vydání také zakládají nový referent. Současně je zajímavé poznamenat, že norma systému ISBN neobsahuje definici pojmu kniha, neboť se na ní účastníci systému dosud neshodli (107).

Metadata obsažená v registru centrálního subsystému (identifikátorová metadata) jsou důležitým prvkem identifikačního systému, zejména popisující identifikovaný objekt. Podle Hakaly by například identifikační systém ISSN byl k ničemu bez přidaného katalogizačního záznamu (108). Systém Handle vyžaduje spíše technická metadata vztahující se k provozu. Systém DOI má vlastní komplexní metadatové schéma pro popis identifikovaného objektu (102). Systémy URN:NBN nevyžadují podrobnější popisné údaje o referentu, s výjimkou českého systému (3, s. 46).

## **9.2 Požadavky na perzistentní identifikaci digitalizátů knih**

Požadavky na perzistentní identifikaci digitalizátů knih vycházejí z požadavků na vyjádření atributů identity digitalizátu knihy (požadavky „P-2.1“), které byly stanoveny v doporučeném postupu uvedeném v oddíle 8.2. Součástí těchto požadavků je nutnost uvést perzistentní identifikátor vydání tištěné knihy, která je předlohou digitalizátu (P.2.1.a.2.1), a perzistentní identifikátor digitalizátu knihy (P.2.1.b.1). To předpokládá existenci systému pro perzistentní identifikaci vydání tištěné knihy (požadavek „P-0.1“) a systému pro perzistentní identifikaci digitalizátů knih (požadavek „P-0.2“) před zahájením digitalizace knih. Z hlediska identifikačních systémů existují tři možnosti: plně využít stávající identifikační systém, vytvořit nový identifikační systém nebo vytvořit vlastní podsystém existujícího systému. Poslední dvě možnosti připadají v úvahu pouze pro velké centrální organizace (např. národní knihovny).

### 9.2.1 Požadavky pro perzistentní identifikaci vydání tištěných knih

PID-K.1	Užít systém kódů národní bibliografie
PID-K.2	Doplňkově využívat další perzistentní identifikátory vydání
PID-K.3	Doplnit do katalogizačních záznamů „československý kód vydání“

#### Užít systém kódů národní bibliografie (PID-K.1)

Systém ISBN neumožňuje přidělovat identifikátory ISBN zpětně, tj. před zavedením systému v dané zemi (systém ISSN ano). Z tohoto důvodu jej nelze užít pro perzistentní identifikaci všech knih. Vhodným řešením je proto zavést systém kódů národní bibliografie (*national bibliographic number*, NBN), pro které je vyčleněno pole 015 ve standardu MARC a které mohou národní knihovny přidělovat záznamům své národní bibliografie (109) (103). Dané národní knihovny si však samy musejí vytvořit vlastní identifikační systém, který kódy NBN přiděluje (tj. určit jejich syntaxi, pravidla užívání apod.). Kódy NBN jsou již desítky let užívány například ve finské národní knihovně (103). V českém prostředí se však tyto kódy až donedávna neužívaly. Jako vhodné řešení způsobu perzistentní identifikace jsem tedy navrhl jejich zavedení v NK ČR.

K praktické realizaci tohoto návrhu došlo v roce 2010, kdy NK ČR zavedla kódy NBN pod názvem číslo České národní bibliografie (čČNB) a přidělila je všem záznamům české národní bibliografie, a to i zpětně (110). Pouhé zanesení čČNB do záznamu by však nebylo dostatečné. Problém je existence duplicitních záznamů – v případě sloučení hrozí, že kód NBN zrušeného záznamu bude smazán, a tedy nebude zaručena perzistentní identifikace. Jediné schůdné řešení problému, které jsem také navrhl, bylo, aby se zrušené nebo neplatné kódy čČNB v záznamu zachovávaly (v podpoli „\$z“ pole 015 standardu MARC, které je k tomu účelu určené a je opakovatelné). Tento návrh zpočátku vzbudil jistou nevoli (vzhledem k tomu, že záznam může po deduplikacích obsahovat velký počet zrušených kódů čČNB), ale nakonec se přistoupilo k jeho realizaci. Ačkoliv nejde o elegantní řešení, je to jediný možný mechanismus v kontextu problémů deduplikace. NK ČR deklaruje jedinečnost a trvalost tohoto kódu a za tímto účelem také stanovila pravidla pro zacházení s tímto kódem v souborném katalogu (110).

### **Doplňkově využívat další perzistentní identifikátory vydání (PID-K.2)**

Tento požadavek znamená doplňkově užít jiné perzistentní identifikátory, které nelze užít pro všechny knihy. V tomto ohledu musí být užít alespoň identifikátor ISBN pro ty knihy, které jej mají přidělený. V případě identifikátoru ISBN je výhoda, že je současně uvedeno v knize. Kódy NBN nejsou (jde o kódy, které existují pouze v záznamech katalogu). Pokud by byly uvedeny ve všech knihách téhož vydání, nebylo by potřeba řešit problém deduplikace záznamů. Reálně však není možné zajistit, aby do všech těchto knih uchovávaných v knihovnách byly kódy NBN zaneseny (např. vytištěny razítkem).

### **Doplnit do katalogizačních záznamů „československý kód vydání“ (PID-K.3)**

Pro perzistentní identifikaci československých knih vydaných v druhé polovině 20. století je možno užít „československý kód vydání“, který byl těmto knihám přidělován a je uveden v jejich tiráži, případně i na přebalu. Systém přidělování tohoto kódu byl do jisté míry analogický k (později zavedenému) systému ISBN. Tento „československý kód“ se skládal ze tří částí tvořených čísly a oddělených pomlčkou (111, s. 57). První část tvořilo dvojčíferné číslo nakladatele, druhou část pořadové číslo knihy v daném nakladatelství a třetí dvojčíslí označovalo letopočet vydání. Tento identifikátor sice neměl kontrolní součet jako identifikátory systému ISBN, ale přesto je vzácný v tom smyslu, že je uveden přímo v knize. Doporučením pro české knihovny je tedy doplňovat tento identifikátor do katalogizačních záznamů jako další perzistentní identifikátor vydání, a pak jej užít při digitalizaci.

### 9.2.2 Požadavky pro perzistentní identifikaci digitalizátů knih

PID-D.1	Stanovit model referentu
PID-D.2	Stanovit strukturu identifikátoru
PID-D.3	Stanovit identifikátorová metadata
PID-D.4	Udržovat údaje o registrátorovi
PID-D.5	Stanovit pravidla pro řízení životního cyklu identifikátorů
PID-D.6	Stanovit pravidla pro řešení oprav užitím perzistentního identifikátoru
PID-D.7	Stanovit možnosti dereference
PID-D.8	Zajistit technologické a organizační řešení systému perzistentní identifikace

#### **Stanovit model referentu (PID-D-1)**

Stanovit model referentu je rozhodným úvodním prvkem každého identifikačního systému. Referentem musí být digitalizát knihy jako intelektuální entita. Ta musí být dlouhodobě uchovávána bez ohledu na transformace objektu CDO (objekt CDO tedy nemůže být referentem identifikátoru). Základní odlišovacím faktorem musejí být z hlediska doporučeného postupu oddílu 8.3 následující atributy identity: vlastník digitalizátu (P-2.1.b.2), digitalizační projekt (P-2.1.b.6) a datum vzniku digitalizátu (P-2.1.b.5). To znamená, že pokud tutéž knihu digitalizuje jiná organizace nebo jiný digitalizační projekt (jiné nebo stejné organizace), jde o jiný digitalizát knihy a musí dostat jiný perzistentní identifikátor. Datum vzniku musí být zohledněno také, a to z toho důvodu, že v rámci jednoho digitalizačního projektu jedné organizace může dojít k tomu, že tatáž kniha bude zdigitalizována vícekrát (například vinou chybně přiděleného perzistentního identifikátoru vydání). V případě nalezení této duplicity je možné ji buď zachovat, nebo odstranit (podle postupu udaného v požadavku „PID-D.6“). Poslední pravidlem modelu (obsaženém v odlišovacím faktoru uvedeném výše) je, že identifikátor se smí přidělovat pouze vlastníkov digitalizátu.

#### **Stanovit strukturu identifikátoru (PID-D.2)**

Struktura identifikátoru musí být tvořena podle modelu „A-B-C“, v němž „A“ je kód identifikující vlastníka digitalizátu, „B“ je kód identifikující digitalizační projekt, „C“ je vlastní část identifikátoru a pomlčka představuje jakékoliv vhodné odlišovací znaménko.



Všechny tyto části identifikátoru musejí být co nejkratší a obsahovat pouze čísla. Takováto struktura umožňuje na základě identifikátoru digitalizátu zjistit jeho vlastníka i projekt. To je vhodné z řady důvodů, mj. těch, že daný registrátor může perzistentní identifikátory užívat špatně a zjištění této skutečnosti pak může čtenáři posoudit důvěryhodnost jím užívaných identifikátorů a také odlišit odpovědnost registrační autority a registrátora.

### **Stanovit identifikátorová metadata (PID-D.3)**

Tento požadavek znamená povinnost registrátora dodávat do identifikačního systému údaje o referentu. To umožňuje, aby identifikační systém sám o sobě mohl poskytovat určité údaje o autenticitě. Identifikátorová metadata musejí zachycovat všechny údaje o attributech identity uvedené v požadavku „P-2.1“ v oddíle 8.3, které lze zaznamenat do metadat. Ostatní údaje musejí být zaznamenány v digitalizační dokumentaci projektu. Z identifikátorových metadat pak musí být možné na základě údaje o názvu digitalizačního projektu tuto dokumentaci získat. Atributy typové digitální identity (P-2.1.b.4) jsou údaje, které se budou v průběhu času měnit. Identifikační systém však musí zachovávat tyto původní údaje, jako typ provenienčních informací. Nutnou součástí identifikátorových metadat musí být samozřejmě URL adresa. Identifikační systém musí uchovávat všechny URL adresy, včetně zaniklých.

### **Udržovat údaje o registrátorovi (PID-D.4)**

Tento požadavek znamená, že identifikační systém musí evidovat o registrátorovi a projektu. Každému registrátorovi a projektu musí být přiřazen jedinečný kód (tyto kódy jsou pak obsaženy ve struktuře identifikátoru (viz požadavek „PID-2“). Tyto údaje musejí obsahovat název organizace a název projektu a případně další identifikační údaje, které lze odvodit z daných legislativních a projektových požadavků na identifikaci organizace a projektu. Identifikační systém musí udržovat číselník kódů vlastníků a kódů projektu. Identifikační systém musí odkazovat na digitalizační dokumentaci, optimálně by ji však měl získat a uchovávat spolu s údaji o projektu ve svém centrálním subsystému.

### **Stanovit pravidla pro řízení životního cyklu identifikátorů (PID-D.5)**

Tento požadavek lze rozdělit do tří oblastí: produkce, archivace a zpřístupňování

Registrátor se musí evidovat v identifikačním systému a dodat mu požadované údaje o organizaci a digitalizačním projektu. Při produkci musí být identifikátor přidělen až ve fázi vytváření balíčku SIP, kdy již existuje digitální originál a kompletují se metadata. Registrátor jej musí uvést do metadat k objektu CDO. Registrátor ve stejné době také musí výměnou za přidělení identifikátoru dodat do identifikačního systému identifikátorová metadata.

Při archivaci musí být identifikátor udržován napříč všemi verzemi balíčku AIP.

Při zpřístupňování musí být identifikátor udržován ve všech balíčcích DIP a při prezentaci v digitální knihovně zobrazován spolu s popisnými metadaty digitalizátu knih tak, aby byl čtenářem ihned registrován. V digitální knihovně musí být uveden odkaz na resolver identifikačního systému spolu s informacemi o tom, jak lze perzistentní identifikátory v tomto resolveru užívat. Digitální knihovna musí do identifikačního systému dodávat informace o aktuální adrese URL (například formou vystavení vazby mezi URL a souvisejícím perzistentním identifikátorem ve svém metadatovém repozitáři OAI, který bude sklízet evidenční služba identifikačního systému).

#### **Stanovit pravidla pro řešení oprav užitím perzistentního identifikátoru (PID-D.6)**

Opravy digitalizátů knih musejí probíhat tak, že bude vytvořen nový dokument, kterému bude přidělen nový identifikátor, přičemž registrátor musí do identifikačního systému dodat informaci o identifikátoru zaniklého dokumentu a údaj o důvodu zániku původního referentu.

#### **Stanovit možnosti dereference (PID-D.7)**

Tento požadavek znamená, že resolver musí primárně provádět přesměrovávání na aktuální adresu. V případě, že je digitalizát vystaven ve více digitálních knihovnách (běžný jev v knihovnické praxi), pak musí přesměrovávat na URL adresu digitální knihovny registrátora. Dále musí mít uživatel možnost v resolveru volby typu dereference (přesměrování, nebo zobrazení identifikátorových metadat). Doplnkově může identifikační systém provádět dereferenci URL adresy (to znamená, že v případě zániku původní adresy bude moci čtenář zjistit, jaký dokument na ní byl vystaven, případně se dostat k jeho perzistentnímu identifikátoru, na základě kterého může najít novou URL adresu).

V případě zániku digitalizátu knihy (opravou) musí být výchozím nastavením dereference zobrazení identifikátorových metadat s odkazem na nový perzistentní identifikátor dokumentu, který vznikl opravou. Splnění těch požadavků výrazně přispívá k citační integritě. Čtenář, který citoval zaniklý dokument, bude užitím resolveru přesměrován na popisnou stránku resolveru, kde bude uvedeno, že dokument zanikl, typ změny, která byla provedena (který je současně důvodem zániku) a identifikátor nového dokumentu, který je označuje nový dokument. Teprve po kliknutí na tento nový identifikátor bude uživateli nabídnuta možnost přesměrovávání.

Tato pravidla umožňují: a) jasně a nepřehlédnutelně upozornit na změnu dokumentu, b) zjistit důvod zániku, c) ukázat vazbu mezi zaniklým a novým dokumentem a umožnit přesměrovávání na nový, d) v případě, že dokument zcela zanikne nebo bude již nepřístupný, se čtenář díky identifikačnímu systému vždy alespoň dozví, že daný dokument existoval a základní informace o něm.

### **Zajistit technologické a organizační řešení systému perzistentní identifikace (PID-D.8)**

Tento požadavek v první řadě znamená, že znamená rozhodnutí, zda budovat vlastní identifikační systém, který bude splňovat předchozí požadavky (PID-D.1 až PID-D.7), či využije stávající. Vzhledem k tomu, že systémy Handle a DOI nejsou vhodné pro identifikaci digitalizátů knih (jeden z důvodů finanční nákladnost, jeden z důvodů, že ne všechny uvedené požadavky mohou naplňovat), vhodnou variantou je pouze systém URN:NBN, který není zpoplatněn a jehož pravidla si může národní knihovna do značné míry stanovovat sama. Všechny výše uvedené požadavky mohou taková konkrétní pravidla národního systému URN:NBN zahrnovat. Tyto důvody mě vedly k tomu, že jsem vytvořil návrh systému ČIDLO, který je českou implementací systému URN:NBN. V systému ČIDLO je většina uvedených požadavků zohledněna, některé se však nepodařilo prosadit, resp. jsou výsledkem mé pozdější výzkumné práce.

Dalšími technickými požadavky této sekce jsou: schopnost digitalizačního systému komunikovat s centrálním subsystémem registrační autority, generovat identifikátory vždy centrálním subsystémem a schopnost digitální knihovny umožnit pravidelné sklizení URL adres. Organizačními požadavky jsou dodržování všech uvedených pravidel všemi účastníky systému, včetně povinnosti registrátora dodávat do identifikačního systému identifikátorová metadata.

## 10 Závěr

Problematika autenticity digitálních dokumentů představuje nepochybně jednu z důležitých oblastí informační vědy. Tato práce vycházela z přejaté teze, že pro specifický typ dokumentu je potřeba stanovit specifický okruh požadavků na udržení a zachování autenticity. Tato teze byla také touto prací potvrzena, jak je zřejmé na příkladech celé řady specifických problémů autenticity digitalizátů knih, které nelze obecně aplikovat na všechny typy dokumentů. Digitalizované knihy představují jeden z hlavních typů dokumentů digitálního dědictví, který je v praxi často vytvářen, uchováván i zpřístupňován stejnou organizací. V této oblasti však dosud neexistuje soustavná konceptualizace autenticity digitalizátů knih v celém jejich životním cyklu, což bylo také hlavním opodstatněním zpracování disertační práce na toto téma.

Při analýze konceptualizací a dílčích tezích o autenticitě ve zdrojích z oblasti digitálního dědictví bylo zjištěno, že různé jimi popisované nebo implicitně obsažené prvky a aspekty autenticity lze zobecnit do následujícího konceptu:

- autenticita vždy předpokládá časoprostorovou distanci daného aktuálního objektu (tj. toho, jehož autenticita se posuzuje) k nějakému referenčnímu objektu, z něhož byl aktuální objekt odvozen řetězcem odvozování;
- udržení autenticity znamená, že v tomto procesu nedošlo ke změně klíčových vlastností digitálního objektu;
- možnost posouzení autenticity vyžaduje, aby byla vytvářena a udržována dokumentace přidružená k objektu popisující 1) vlastní intelektuální obsah digitálního objektu, 2) celou historii digitálního objektu (jeho původ a vznik a řetězec odvozování a správy).

Požadavek dokumentace je specifickým požadavkem autenticity v kontextu institucionální správy dokumentů a znamená nutnost systematického vytváření a udržování dokumentace spravující organizací.

V disertační práci byl vytvořen obecnější rámec pro autenticitu, jehož výchozím zdrojem je model OAIS, a který je dále specifikován převzetím, úpravou a rozšířením modelu autenticity projektu InterPARES (autenticita jako celek dvou složek dokumentu: jeho identity a integrity) následujícím způsobem. Z pohledu modelu OAIS je autenticita především autenticitou informačního obsahu, který je tvořen datovým objektem s obsahem (objekt

CDO), který tento obsah nese, a interpretačními informacemi (které tvoří informace, jakým způsobem je informační obsah zakódován v objektu CDO, což zahrnuje zejména informace o jeho formátu a softwaru pro jeho adekvátní reprodukci), na základě kterých může být objekt CDO převeden do podoby informací srozumitelných člověku (tedy informačního obsahu jakožto intelektuální entity). Životní cyklus informačního obsahu lze chápat jako posloupnost odvozování (tj. transformací a kopírování) objektu CDO a jeho adekvátní aktuální reprodukce do podoby intelektuální entity vnímatelné člověkem. Identita intelektuální entity (reprodukováného informačního obsahu) odpovídá konceptu identifikačních informací a části provenienčních informací (těch, které popisují původ a zdroj informačního obsahu), a její integrita odpovídá adekvátním interpretačním informacím, které dokáží informační obsah reprodukovat způsobem, který je v souladu s klíčovými vlastnostmi intelektuální entity, což v kontextu digitálních knihoven znamená především výběr vhodného zpřístupňovacího softwaru a jeho udržování v závislosti na aktuálních verzích internetového prohlížeče. Identita a integrita objektu CDO znamená udržování takové digitální identity, která je v souladu s klíčovými vlastnostmi intelektuální entity a která umožňuje intelektuální entitu kdykoliv reprodukovat. Digitální identita se může měnit (např. formátovou konverzí se nutně mění formát), ale klíčové vlastnosti intelektuální entity musejí být zachovány. Pro udržení integrity objektu CDO jsou potřebné informace, které odpovídají konceptu informací o neporušenosti a části provenienčních informací (těch, které popisují historii odvozování) – tyto typy dodatečných informací musejí být podle modelu OAIS uchovávány spolu s informačním obsahem.

Tento obecný rámec byl dále aplikován při analýze životního cyklu digitalizátů knih, jejímž výsledkem je konceptualizace požadavků autenticity pro tento typ dokumentu. Za hlavní problémové okruhy byly určeny:

- proceduralita (nutnost dodržovat způsob odvozování informačních balíčků, a tím i informačního obsahu v nich uloženého, podle logiky modelu OAIS);
- identita a identifikace (specifikovaná na mnohem širší a detailnější úrovni než v modelu projektu InterPARES, přičemž rozhodujícím prvkem je vymezení referenčního objektu a perzistence identifikace intelektuální entity);
- validace digitální identity (zejména formátu) jako specifická a nebytná forma ověření integrity v počítačovém prostředí;
- dostupnost a důvěryhodnost nástrojů (závislost na kvalitních nástrojích a problémy jejich vývoje, podpory a dokumentace);

- úplnost a rozsah dokumentace;
- komplexita digitální identity (zejména formátů a obrazových specifikací);
- opravy digitalizovaných knih.

Posledně jmenovaný problém byl označen za patrně nejpalčivější problém autenticity digitálních dokumentů, který ohrožuje využití informací uživateli, zejména citační praxi (uživateli je v průběhu času nabízena nejprve vadná podoba dokumentu a následně opravená, výrazně se lišící, ale užita a citována byla prvně jmenovaná).

Specifikem digitálních dokumentů obecně je skutečnost, že pro zachování a udržení jejich autenticity jsou nepostradatelnými prvky technologický rámec a metadatové standardy. Technologický rámec determinuje vyhlídky zachování autenticity. V případě digitalizátů knih jsou tyto vyhlídky pozitivní v tom smyslu, že existují zejména vhodné formáty (v jiných oblastech tomu tak vždy není) a jejich validátory. Komplexita digitální identity však znamená, že například ne každý nástroj dokáže vytvořit autentický formát (tj. takový, který je plně v souladu se svou oficiální specifikací – objekt ve vadně vytvořeném formátu se tedy vydává za něco jiného, než čím je); různé verze validátoru mohou přinášet rozdílné výsledky; různé verze nástroje pro identifikaci mohou odlišně určit formát; objekt ve formátu, který nemá otevřenou specifikaci, nelze transparentně validovat. Bylo konstatováno, že vzhledem k tomu, že jediným dostupným způsobem jednoznačné identifikace formátu je identifikátor PUID, který je svázán s registrem formátů PRONOM, je nutno doporučit, aby se organizace v prvé řadě spoléhaly na tento způsob identifikace a využívali nástroj DROID, který je vyvíjen stejnou organizací, a záruka jeho důvěryhodnosti je tak nejvyšší (a případně se pokusily přispět k podpoře této služby).

Rozumným přístupem v oblasti metadat je využívat pouze široce rozšířené a otevřené metadatové standardy; pro digitalizáty knih jsou klíčové zejména standardy METS, PREMIS a MIX. Je současně potřeba mít na zřeteli, že také tyto standardy vymezují pole možného (co jimi lze a nelze zapsat a v jaké logice). V současnosti například neexistuje podrobné a rozšířené schéma pro zápis klíčových vlastností digitalizátu knih – ve standardu PREMIS (kde tomuto konceptu odpovídá sekce „significant properties“) je pouze možno libovolně definovat vlastní typ vlastnosti a její hodnotu, ale způsob konkrétní implementace je obtížný vzhledem k vysoké komplexitě takové definice, která musí být dostatečně obecná a umožňovat jednoznačnou verifikaci přítomnosti takto stanovené vlastnosti v objektu.

Základem doporučeného postupu, který tato disertační práce navrhuje, je přesná a komplexní definice referenčního objektu. V normě OAIS konceptu referenčního objektu odpovídá objekt CDO a jeho interpretační informace uložené v první verzi archivního informačního balíčku. Pro řízení celého životního cyklu digitalizovaného dokumentu jednou organizací je však nutno tuto definici vytvořit již před zahájením digitalizace – jako cíl digitalizace i následného dlouhodobého uchovávání, a to vzhledem k vazbě digitalizátu na tištěnou předlohu. Digitalizovat knihu lze různými způsoby – například formou věrné barevné digitalizace, černobílé digitalizace nebo pouhým přepisem textu. Nejčastějším typem digitalizace knih, který byl hlavním předmětem zkoumání, je věrná digitalizace do podoby rastrových dat. Doporučený postup se tak specificky blíže zaměřuje na tento případ. Byly identifikovány tři komponenty digitalizátu knihy: obrazová komponenta, textová data (jako výstup technologie optického rozpoznávání znaků v obrazových datech, umožňující plnotextové prohledávání) a strukturální komponenty. Před zahájením digitalizace je nutné stanovit a vyjádřit atributy identity intelektuální entity:

- popsat exemplář knihy a časové určení (současný stav / předpokládaný původní stav) a podle toho uvést metodu převodu (restaurování, věrné zachycení aktuálního stavu);
- explicitně stanovit úroveň abstrakce podle FRBR (tedy exemplář a vydání, s uvedením, na jakou úroveň je kladen důraz), která je předmětem snímání;
- specifikovat klíčové vlastnosti intelektuální entity, které bude mít digitální originál a které je nutno dlouhodobě uchovávat;
- podrobně popsat postup odvozování tištěné knihy do podoby digitálního originálu;

Z hlediska identity intelektuální entity byly za nezbytné informace určeny:

- identifikátor exempláře;
- informace o stavu degradace a barevných vlastnostech předlohy a jeho vlastníkovi;
- perzistentní identifikátor vydání;
- perzistentní identifikátor digitalizátu knih;
- popis klíčových vlastností intelektuální entity (např. barevná věrnost, plnotextová prohledatelnost);

- název digitalizačního projektu a jeho dokumentace;
- název vlastníka digitalizátu;
- údaj o datu vzniku digitalizátu;

Z těch informací, které se vztahují k typovému popisu, je zásadně důležité stanovení a dokumentace klíčových vlastností intelektuální entity, což je řešení, jak nahradit v současnosti nedostupné podrobné metadatové schéma pro zápis klíčových vlastností, a současně způsobem, jak čtenářské obci srozumitelným způsobem popsat, jaké vlastnosti lze od prezentovaného digitalizátu očekávat. Čtenáři by měly být dostupné všechny tyto informace, neboť určují referenční objekt, na základě kterého lze autenticitu posuzovat. Čtenář například nesmí být uveden v omyl, že dokument prezentovaný v digitální knihovně je nativně digitální kniha (e-born publikace), a nikoliv digitalizovaná kniha, a musí mít možnost na základě digitalizátu dohledat tištěnou předlohu.

Pro integritu objektu CDO je důležité, aby byly při všech odvozováních (zejména při formátových konverzích) zachovány klíčové vlastnosti intelektuální entity. Ty vlastnosti je nutno pro každou transformaci konkrétně specifikovat do podoby digitální identity, která má být cílem transformace (např. matematicky bezztrátová komprese implikuje zachování barevné věrnosti). Bez ohledu na potíže s nedostupností komplexního metadatového schématu pro klíčové vlastnosti je třeba zdůraznit, že není možné zobecnit všechny technické informace, které mají být zachovány. Za klíčový atribut digitální identity současnosti lze považovat například barevný profil Adobe RGB 1999, současně však neexistuje způsob, jak zobecnit obrazovou kvalitu prezentovanou tímto prostorem tak, aby v budoucnosti, až tento prostor zanikne a bude nahrazen jiným, bylo možno na základě této obecné definice určit, jaký bude jeho vhodný nástupnický barevný prostor. Navrhovaným řešením je proto definovat některé klíčové technické vlastnosti konkrétně (např. uvést název tohoto profilu) a v dokumentaci archivu zaznamenat požadavek, že pokud v budoucnosti barevný prostor, který tento profil specifikuje, začne zastarávat, archiv provede výzkum a nahradí tento údaj názvem profilu, který určí za nástupnický a který bude nadále zachováván (do doby, než zase zastará, a celý proces se tak bude opakovat).

Posledními z nejdůležitějších bodů doporučeného postupu je komplexní dokumentace, zejména digitalizačního projektu, a to vzhledem k jejímu časově omezenému trvání. Tato dokumentace musí zahrnovat detailní popis způsobu plnění metadatového profilu (jaké položky jakými nástroji jaké verze budou plněny a pomocí jakých řízených slovníků) a popis



výše uvedených voleb. Pokud digitalizační projekt zahrnuje více odlišných postupů, musejí být v rámci dokumentace odlišeny.

Na základě provedené analýzy byla jako klíčová složka pro udržení a ověření autenticity zjištěna potřeba perzistentní identifikace intelektuální entity. Za tímto účelem byl vytvořen návrh pravidel pro systém perzistentní identifikace digitalizátů knih jakožto intelektuální entity. Jádrem návrhu je využití perzistentního identifikátoru jak pro identifikaci intelektuální entity napříč odvozováním objektu CDO (z časového i prostorového hlediska, tj. vazba mezi digitalizátem knihy uloženým v archivním úložišti na jedné straně a zpřístupňovaným v digitální knihovně na straně druhé), tak pro přesměrování na aktuální URL adresu. Součástí (proměnlivé) digitální identity je nepochybně URL adresa dokumentu v internetové síti. Ta se však mění (tak jako se mění objekt CDO). Řešením je trvalé propojování aktuálních URL adres s neměnným perzistentním identifikátorem intelektuální entity identifikačního systému s digitalizačním systémem a archivem (včetně digitální knihovny). Součástí registru identifikačního systému musejí být základní metadata o identifikovaném digitalizátu knih. Z nich mezi nejdůležitější klíčové údaje patří identifikátor vydání tištěné knihy a název digitalizačního projektu (případně identifikace jeho konkrétního postupu, kterým objekt vznikl) a odkaz na dokumentaci tohoto projektu.

Pro propojení mezi digitalizátem knih a tištěnou předlohou je důležité zavedení národního systému pro perzistentní identifikaci vydání knih (identifikátor ISBN nepokrývá celou národní bibliografii). Za tímto účelem lze využít kód národní bibliografie, který je definován standardem MARC. Dále jsem navrhl, aby se v českých knihovnách při katalogizaci začal zaznamenávat i specifický československý identifikátor vydání (přidělovaný v 50.-80. letech), který je uveden v samotné knize, čímž by se usnadnila kontrola bibliografického záznamu. Návrh pravidel pro systém perzistentní identifikace digitalizátů knih však také umožňuje řešení problému oprav digitalizátů knih, které byly v této disertační práci shledány za nejmarkantnější současné riziko narušení autenticity. V případě výrazných změn digitalizátu knihy (rámec pro posuzování takových změn je součástí návrhu pravidel perzistentní identifikace) musí být původní dokument smazán z archivního úložiště i digitální knihovny a musí být vytvořen nový (opravou nebo opětovnou digitalizací, kterou doporučujeme ve všech případech, kdy to možné je), kterému bude přidělen nový identifikátor se současným uvedením vazby na původní identifikátor. Dereference původního dokumentu v resolveru pak bude zahrnovat pouze výpis metadat z registru identifikačního systému, které budou popisovat původní dokument, důvod jeho zániku (typ chyby / opravy) a odkaz na nový identifikátor. Teprve z nového identifikátoru bude možno provést

přesměrování na nový dokument. Tento mechanismus umožňuje a) zjistit, že citovaný dokument kdysi existoval („autenticita citace“), b) jasně upozornit čtenáře na skutečnost změny dokumentu, a tak snížit riziko jeho uvedení v omyl, resp. narušení důvěryhodnosti jeho citace.

Uvedená pravidla perzistentního systému identifikace pak byla částečně užita při mé implementaci mezinárodního systému URN:NBN pro české prostředí, jejímž výsledkem je systém ČIDLO.<sup>94</sup>

---

<sup>94</sup> <https://resolver.nkp.cz/>

## Seznam použitých pramenů a literatury

1. CUBR, Ladislav. *Dlouhodobá ochrana digitálních dokumentů*. Praha: Národní knihovna ČR, 2010, 154 s. ISBN 978-80-7050-588-5.
2. CUBR, Ladislav a Martin ŘEHÁNEK. *CZIDLO (CZech IDentification and LOcalization Tool): výsledek* [online]. Praha: ČVÚT, 2012 [cit. 2013-07-23]. Dostupné z: <http://aplikace.isvav.cvut.cz/vklap/>.
3. CUBR, Ladislav et al. Srovnání vybraných národních identifikačních systémů užívajících identifikátory URN:NBN. *ProInflow: časopis pro informační vědy* [online]. 2016, **8**(1) [cit. 2017-03-10]. Dostupné z: <http://www.phil.muni.cz/journals/index.php/proinflow/article/view/1220>.
4. CUBR, Ladislav. Budování důvěryhodného systému trvalé identifikace digitálních dokumentů. *Knihovna* [online]. 2010, **21**(1), s. 23-31.
5. CUBR, Ladislav. *Využití perzistentních identifikátorů pro elektronické publikace*. Praha, 2013.
6. CUBR, Ladislav a Zdeněk VAŠEK. Identifikátory digitálních dokumentů se zaměřením na systém URN:NBN v ČR. *Čtenář: měsíčník pro knihovny*. 2013, **65**(6), s. 214-217.
7. VAŠEK, Zdeněk, Ladislav CUBR a KOL. Metodika pro přidělování a správu životního cyklu unikátních perzistentních identifikátorů digitálních dokumentů podle standardu URN:NBN. In: *Národní digitální knihovna* [online]. Praha: Národní knihovna ČR, [2015 cit. 2017-03-10]. URN: urn:nbn:cz:nk-0027gj. Dostupné z: [http://www.ndk.cz/archivace/Metodika\\_URNNBN\\_final\\_2.1a.pdf](http://www.ndk.cz/archivace/Metodika_URNNBN_final_2.1a.pdf).
8. CAPLAN, Priscilla. What Is Digital Preservation? *Library Technology Reports*. Feb-Mar 2008, **44**(2), s. 7-9.
9. ISO 14721:2003. *Space data and information transfer systems - Open archival information system (OAIS) - Reference model*. Geneva: ISO, 2003, iii, 156 s.
10. ISO 14721:2012. *Space data and information transfer systems - Open archival information system (OAIS) - Reference model*. 2nd ed. Geneva: ISO, 2012, 126 s.
11. ISO 16363:2012. *Space data and information transfer systems - Audit and certification of trustworthy digital repositories*. Geneva: ISO, 2012, 70 s.
12. Consultative Committee for Space Data Systems. *Reference Model for an Open Archival Information System (OAIS)* [online]. In: *Blue Book*. Washington (DC): CCSDS Secretariat, January 2002, issue 1 [cit. 2015-03-07]. CCSDS 650.0-B-1. Dostupné z: <http://public.ccsds.org/publications/archive/650x0b1s.pdf>.
13. Consultative Committee for Space Data Systems. *Reference Model for an Open Archival Information System (OAIS): Recommended Practice* [online]. In: *Blue Book*. Washington (DC): CCSDS Secretariat, June 2012, issue 2 [cit. 2017-03-14]. CCSDS 650.0-M-2. Dostupné z: <https://public.ccsds.org/pubs/650x0m2.pdf>.
14. CENTER FOR RESEARCH LIBRARIES a OCLC. *Trustworthy Repositories Audit & Certification: Criteria and Checklist* [online]. Version 1.0. Chicago: CRL, February 2007, 88 s. [cit. 2017-03-27]. Dostupné z: [http://www.crl.edu/sites/default/files/d6/attachments/pages/trac\\_0.pdf](http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf).

15. Consultative Committee for Space Data Systems. *Audit and Certification of Trustworthy Digital Repositories: Recommended Practice* [online]. In: *Magenta Book*. Washington (DC): CCSDS Secretariat, Sept 2011, issue 1 [cit. 2017-03-14]. CCSDS 652.0-M-1. Dostupné z: <https://public.ccsds.org/pubs/652x0m1.pdf>.
16. ČSN ISO 14721. *Systémy pro přenos dat a informací z kosmického prostoru - Otevřený archivační informační systém - Referenční model*. 2. vyd. Praha: Úřad pro technickou normalizaci, metrologii a státní zkušebnictví, 2014, 97 s. Třídící znak 31 9620.
17. ČSN ISO 16363. *Systémy pro přenos dat a informací z kosmického prostoru – Audit a certifikace důvěryhodných digitálních úložišť*. Praha: Úřad pro technickou normalizaci, metrologii a státní zkušebnictví, 2014. 53 s. Třídící znak 31 9621.
18. VICKERY, B. C. *Information systems*. London: Butterworths, 1973, 350 s. ISBN 04-087-0456-X.
19. DURANTI, Luciana a Kenneth THIBODEAU. The concept of record in interactive, experiential and dynamic environments: the view of InterPARES. *Archival Science*. 2006, 6(1), s. 13 - 68.
20. FEDERAL AGENCIES DIGITIZATION INITIATIVE, Still Image Working Group. *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files* [online]. Washington (DC): FADGI, Aug 2010, 96 s. [cit. 2017-03-15]. Dostupné z: [http://www.digitizationguidelines.gov/guidelines/FADGI\\_Still\\_Image-Tech\\_Guidelines\\_2010-08-24.pdf](http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf).
21. THALLER, Manfred. Characterisation. In: *Digital Preservation Planning: Principles, Examples and the Future with Planets* [online prezentace]. London: British Library, 29th July 2008 [cit. 2017-03-16]. Dostupné z: [http://www.planets-project.eu/events/bl\\_planning/abstracts/manfred\\_thaller.pdf](http://www.planets-project.eu/events/bl_planning/abstracts/manfred_thaller.pdf).
22. LAWRENCE, Gregory W. et al. *Risk management of digital information: a file format investigation*. Washington (DC): Council on Library and Information Resources, c2000, vi, 75 s. ISBN 18-873-3478-5. Dostupné také z: <https://www.clir.org/pubs/reports/pub93/pub93.pdf>.
23. NIELSEN, Anders Bo a Alex THIRIFAYS. Cost Aspects of Ingest and Normalization. In: BORBINHA, José et al., ed. *iPRES 2011: 8th International Conference on Preservation of Digital Objects* [online]. Singapore: National Library Board, 2011, s. 107-115 [cit. 2017-03-14]. ISBN 978-981-07-0441-4. Dostupné z: [https://phaidra.univie.ac.at/detail\\_object/o:294293](https://phaidra.univie.ac.at/detail_object/o:294293).
24. CHAPMAN, Stephen. *Page Image Compression for Mass Digitization* [online]. Cambridge (MA): Harvard Library, 2006, 6 s. [cit. 2017-03-27]. Dostupné z: [http://library.harvard.edu/sites/default/files/images/IST\\_PageImageCompression\\_preprint.pdf](http://library.harvard.edu/sites/default/files/images/IST_PageImageCompression_preprint.pdf). Published in Archiving 2007: Final Program and Proceedings.
25. Sustainability of Digital Formats: Planning for Library of Congress Collections. *Digital Preservation (Library of Congress)* [online]. Washington (DC): Library of Congress, Last updated 07/24/2013 [cit. 2017-03-16]. Dostupné z: <http://www.digitalpreservation.gov/formats/index.shtml>.
26. LIBRARY OF CONGRESS. *Recommended Formats Statement 2016-2017* [online]. Washington (DC): The Library of Congress, [2017 cit. 2017-03-01]. Dostupné z: <http://www.loc.gov/preservation/resources/rfs/RFS%202016-2017.pdf>.
27. Raster Still Images for Digitization: A Comparison of File Formats. *Federal Agencies Digital Guidelines Initiative* [online]. Washington (DC): FADGI, Last Updated: 9/9/2014 [cit. 2017-03-16]. Dostupné z: [http://www.digitizationguidelines.gov/guidelines/raster\\_stillImage\\_compare.html](http://www.digitizationguidelines.gov/guidelines/raster_stillImage_compare.html).

28. FERNIE, Kate, ed. *Technical Guidelines for Digital Cultural Content Creation Programmes*. Version 2.0. [Rome]: Minerva EC, 2008, 92 s. Dostupné také z: <http://www.minervaeurope.org/publications/MINERVA%20TG%202.0.pdf>.
29. ISO/IEC 15444-1:2004. *Information technology - JPEG 2000 image coding system: core coding system*. 2nd ed. Geneva: ISO, 2004.
30. NISO FRAMEWORK WORKING GROUP. *A framework of guidance for building good digital collections: a NISO recommended practice* [online]. 3rd ed. Baltimore (MD): National Information Standards Organization (NISO), 2007, iii, 95 s. [cit. 2017-03-15]. ISBN 978-1-880124-74-1. Dostupné z: <http://www.niso.org/publications/rp/framework3.pdf>.
31. RIMKUS, Kyle, Thomas PADILLA, Tracy POPP a Greer MARTIN. Digital Preservation File Format Policies of ARL Member Libraries: An Analysis. *D-Lib Magazine* [online]. 2014, **20**(3/4) [cit. 2017-03-13]. DOI: 10.1045/march2014-rimkus. Dostupné z: <http://www.dlib.org/dlib/march14/rimkus/03rimkus.html>.
32. FEDERAL AGENCIES DIGITIZATION INITIATIVE, Still Image Working Group. *Raster Still Images for Digitization: A Comparison of File Formats* [online]. Part 3. Narrative and Summary Table. Washington (DC): FADGI, Revised, Aug 29, 2014, 9 s. [cit. 2017-03-15]. Dostupné z: [http://www.digitizationguidelines.gov/guidelines/FADGI\\_RasterFormatCompare\\_p3\\_20140829\\_r.pdf](http://www.digitizationguidelines.gov/guidelines/FADGI_RasterFormatCompare_p3_20140829_r.pdf).
33. SANZ, Pascal. Development of electronic periodicals at the Bibliothèque nationale de France: digitisation of French daily newspapers from Mid 19th Century to 1944. In: *World library and information congress: 71st IFLA General Conference and Council: "Libraries: a voyage of discovery": August 14th-18th 2005, Oslo, Norway* [online]. Hague: IFLA, 2005 [cit. 2017-03-14]. Dostupné z: [http://archive.ifla.org/IV/ifla71/papers/141e\\_trans-Sanz.pdf](http://archive.ifla.org/IV/ifla71/papers/141e_trans-Sanz.pdf).
34. JOHANSSON, Henrik, Per Erik SVEDLUND, Erik SIIRA a Hamid SARVE. A cost-efficient and automatic digitization workflow using commodity hardware and image analysis. In: *Archiving 2010: preservation strategies and imaging technologies for cultural heritage institutions and memory organizations: final program and proceedings, June 1-4, 2010, Koninklijke Bibliotheek, Den Haag, the Netherlands*. Springfield (VA): Society for Imaging Science and Technology, 2010, s. 101-106. ISBN 978-0-89208-290-2.
35. VAN DER KNIJFF, Johan. JPEG 2000 for Long-term Preservation: JP2 as a Preservation Format. *D-Lib Magazine* [online]. 2011, **17**(5/6) [cit. 2017-03-13]. DOI: 10.1045/may2011-vanderknijff. Dostupné z: <http://www.dlib.org/dlib/may11/vanderknijff/05vanderknijff.html>.
36. BUONORA, Paolo a Franco LIBERATI. A Format for Digital Preservation of Images. *D-Lib Magazine* [online]. 2008, **14**(7/8), - [cit. 2017-03-31]. DOI: 10.1045/july2008-buonora. Dostupné z: <http://www.dlib.org/dlib/july08/buonora/07buonora.html>.
37. BUCKLEY, Robert. *JPEG 2000 as a Preservation and Access Format for the Wellcome Trust Digital Library* [online]. London: King's College London, Aug 2009, 17 s. [cit. 2017-03-15]. Dostupné z: <http://wellcomelibrary.org/content/documents/22082/JPEG2000-preservation-format.pdf>.
38. BRYGFJELD, Svein Arne. JP2K for preservation and access, experiences from the National Library of Norway. In: *JPEG 2000 for the Practitioner* [online]. Glasgow: Digital Preservation

- Coalition, 16 Nov 2010 [cit. 2017-03-15]. Dostupné z:  
<http://www.dpconline.org/docman/miscellaneous/events/521-jp2knov2010brygfjeld/file>.
39. Standardy pro obrazová data. *Národní digitální knihovna* [online]. Praha: Národní knihovna ČR, 23. 03. 2015 [cit. 2016-07-19]. Dostupné z: <http://www.ndk.cz/standardy-digitalizace/standardy-pro-obrazova-data>.
40. JPEG 2000 profiles – examples from a range of institutions. In: *JPEG 2000 for the Practitioner* [online]. Glasgow: Digital Preservation Coalition, 16 Nov 2010 [cit. 2017-03-15]. Dostupné z: <http://www.dpconline.org/docman/miscellaneous/events/529-jp2knov2010parametercomparisonchart/file>.
41. CUBR, Ladislav. Formátová strategie LTP úložiště NK ČR. In: *CDA 2016: formátové výzvy LTP: zborník príspevkov z 1. medzinárodnej konferencie o dlhodobej archivácii*. Bratislava: Univerzitná knižnica v Bratislave, 2016, s. 44-57. ISBN 978-80-89303-51-9. ISSN 2453-9406.
42. PRONOM. *The National Archives* [online]. London: National Archives, 2017 [cit. 2017-03-27]. Dostupné z: <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>.
43. BROWN, Adrian. *The PRONOM PUID Scheme: a scheme of persistent unique identifiers for representation information* [online]. London: National Archives, 27 July 2006, 9 s. [cit. 2017-03-21]. Digital Preservation Technical Paper, issue 2. Dostupné z: [http://www.nationalarchives.gov.uk/aboutapps/pronom/pdf/pronom\\_unique\\_identifier\\_scheme.pdf](http://www.nationalarchives.gov.uk/aboutapps/pronom/pdf/pronom_unique_identifier_scheme.pdf). Document Reference: DPTP-02.
44. BUCKLEY, Robert. *JPEG 2000 - a Practical Digital Preservation Standard?* [online]. Glasgow: Digital Preservation Coalition, February 2008, 21 s. [cit. 2017-03-31]. DPC Technology Watch Series Report, 08-01. Dostupné z: <http://www.dpconline.org/docman/technology-watch-reports/87-jpeg-2000-a-practical-digital-preservation-standard/file>.
45. THE ASSOCIATION FOR LIBRARY COLLECTIONS AND TECHNICAL SERVICES, Preservation & Reformatting Section. Minimum Digitization Capture Recommendations. In: *Association for Library Collections & Technical Services (ALCTS)* [online]. Chicago: ALA, June 2013 [cit. 2017-03-14]. Dostupné z: <http://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations>.
46. Color space - Glossary. *Federal Agencies Digital Guidelines Initiative* [online]. Washington (DC): FADGI, 2017 [cit. 2017-03-16]. Dostupné z: <http://www.digitizationguidelines.gov/term.php?term=colorspace>.
47. ALTO Principles. *ALTO: Technical Metadata for Layout and Text Objects (Standards, Library of Congress)* [online]. Washington (DC): Library of Congress, June 8, 2016 [cit. 2017-03-13]. Dostupné z: <https://www.loc.gov/standards/alto/description.html>.
48. ČESKO, Ministerstvo kultury, Odbor umění, literatury a knihoven. Veřejné informační služby knihoven (VISK): podprogram č. 7: národní program ochrany a digitalizace dokumentů ohrožených degradací kyselého papíru - KRAMERIUS. In: *Veřejné informační služby knihoven* [online]. Praha: Národní knihovna ČR, 16. 9. 2015, 8 s. [cit. 2017-03-20]. Dostupné z: <http://visk.nkp.cz/dokumenty/visk7/2016/VISK7-podm2016.doc>.
49. ZENG, Marcia Lei a QIN, Jian. *Metadata*. 2nd edition. Chicago: Neal-Schuman, an imprint of the American Library Association, 2016. xxvii, 555 stran. ISBN 978-1-55570-965-5.
50. METS: An Overview & Tutorial. *Metadata Encoding and Transmission Standard (METS)* [online]. Washington (DC): Library of Congress, February 9, 2016 [cit. 2017-03-13]. Dostupné z: <http://www.loc.gov/standards/mets/METSOverview.v2.html>.

51. *Metadata Encoding and Transmission Standard (METS)* [online]. Washington (DC): Library of Congress, August 9, 2016 [cit. 2017-03-13]. Dostupné z: <http://www.loc.gov/standards/mets>.
52. PREMIS EDITORIAL COMMITTEE. *PREMIS Data Dictionary for Preservation Metadata* [online]. Version 3.0. Washington (DC): Library of Congress, June 2015, rev. Nov 2015, viii, 273 s. [cit. 2016-10-14]. Dostupné z: <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>.
53. MODS: Uses and Features (Metadata Object Description Schema: MODS). *Metadata Object Description Schema: MODS* [online]. Washington (DC): Library of Congress, February 1, 2016 [cit. 2017-03-27]. Dostupné z: <http://www.loc.gov/standards/mods/mods-overview.html>.
54. *Metadata for Images in XML Standard (MIX)* [online]. Washington (DC): Library of Congress, November 23, 2015 [cit. 2017-03-27]. Dostupné z: <http://www.loc.gov/standards/mix/>.
55. DUVAL, Erik, Wayne HODGINS, Stuart SUTTON a Stuart L. WEIBEL. Metadata Principles and Practicalities. *D-Lib Magazine* [online]. 2002, 8(4) [cit. 2017-03-13]. DOI: 10.1045/april2002-weibel. Dostupné z: <http://www.dlib.org/dlib/april02/weibel/04weibel.html>.
56. Standardy pro metadata. *Národní digitální knihovna* [online]. Praha: Národní knihovna ČR, 04. 03. 2016 [cit. 2016-07-18]. Dostupné z: <http://www.ndk.cz/standardy-digitalizace/metadata>.
57. IFLA STUDY GROUP ON THE FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS. Functional Requirements for Bibliographic Records: final report [online]. Haag: IFLA, September 1997, as amended and corrected through Feb 2009, v, 137 s. [cit. 2017-03-27]. Dostupné z: [https://www.ifla.org/files/assets/cataloguing/frbr/frbr\\_2008.pdf](https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf).
58. METS Profiles. *Metadata Encoding and Transmission Standard (METS)* [online]. Washington (DC): Library of Congress, February 9, 2016 [cit. 2017-03-13]. Dostupné z: <http://www.loc.gov/standards/mets/mets-registered-profiles.html>.
59. ANSI/NISO Z39.87-2006. *Data Dictionary – Technical Metadata for Digital Still Images*. Bethesda (MD): NISO Press, 2006, xiv, 107 s. ISBN 978-1-937522-37-7. ISSN 1041-5653.
60. CIPA DC-008-TRANSLATION-2012. *Exchangeable file digital still cameras: exif version 2.3* [online]. Tokyo: CIPA, 2012, 185 s. [cit. 2017-03-14]. Dostupné z: [http://www.cipa.jp/std/documents/e/DC-008-2012\\_E.pdf](http://www.cipa.jp/std/documents/e/DC-008-2012_E.pdf).
61. LIBRARY OF CONGRESS, Office of Strategic Initiatives. *JPEG 2000 Profile for the National Digital Newspaper Program* [online]. Washington (DC): The Library of Congress, April 27, 2006, 24 s. [cit. 2017-02-26]. Dostupné z: <http://www.loc.gov/preservation/resources/rfs/RFS%202016-2017.pdf>.
62. GRYCZ, Czeslaw Jan. Digitising Rare Books and Manuscripts. In: MACDONALD, Lindsay, ed. *Digital heritage: applying digital imaging to cultural heritage*. Amsterdam: Elsevier, 2006, s. 33-68. ISBN 0-75-066183-6.
63. Národní digitální knihovna. *Národní knihovna České republiky* [online]. Praha: Národní knihovna ČR, 27. 05. 2014 [cit. 2017-03-27]. Dostupné z: <https://www.nkp.cz/o-knihovne/zakladni-informace/narodni-digitalni-knihovna>.
64. *OpenJPEG: An open-source JPEG 2000 codec written in C* [online]. Louvain: Université catholique de Louvain (UCL), c2017 [cit. 2017-03-27]. Dostupné z: <http://www.openjpeg.org/>.
65. PALMER, William, Peter MAY a Peter CLIFF. An Analysis of Contemporary JPEG2000 Codecs for Image Format Migration. In: BORBINHA, José, Michael NELSON a Steve KNIGHT, ed. *iPRES 2013: Proceedings / of the 10th International Conference on Preservation of Digital*

- Objects: 3 - 5 September 2013, Lisbon, Portugal* [online]. Lisboa: Biblioteca Nacional de Portugal, 2013, s. 197-202 [cit. 2017-03-14]. ISBN 978-972-565-493-4. Dostupné z: <http://purl.pt/24107/1/>.
66. SMITH, Neil. Digitising Documents for Public Access. In: MACDONALD, Lindsay, ed. *Digital heritage: applying digital imaging to cultural heritage*. Amsterdam: Elsevier, 2006, s. 3-32. ISBN 0-75-066183-6.
  67. VYCHODIL, Bedřich. JPEG2000 - Aneb nemyslete si, že vás mine!. *Knihovna* [online]. 2010, 21(2), s. 53-68.
  68. FEDERAL AGENCIES DIGITIZATION INITIATIVE, Still Image Working Group. *Guidelines for TIFF Metadata: Recommended Elements and Format* [online]. Version 1.0. Washington (DC): FADGI, Feb 10, 2009, 5 s. [cit. 2017-03-15]. Dostupné z: [http://www.digitizationguidelines.gov/guidelines/TIFF\\_Metadata\\_Final.pdf](http://www.digitizationguidelines.gov/guidelines/TIFF_Metadata_Final.pdf).
  69. CARON, Bertrand, Thomas LEDOUX, Stéphane REECHT a Jean-Philippe TRAMONI. Experiment, Document & Decide: a Collaborative Approach to Preservation Planning at the BnF. In: *iPRES 2015: Proceedings of the 12th International Conference on Digital Preservation: Chapel Hill, November 2-6* [online]. Chapel Hill (NC): School of Information and Library Science, 2015, s. 54-58 [cit. 2017-03-14]. ISBN 978-0-692-59881-8. Dostupné z: <https://phaidra.univie.ac.at/view/o:429524>.
  70. Komplexní validátor. *Národní digitální knihovna* [online]. Praha: Národní knihovna ČR, 24. 02. 2017 [cit. 2017-03-22]. Dostupné z: <http://www.ndk.cz/archivace/komplexni-validator>.
  71. DURANTI, Luciana. The impact of digital technology on archival science. *Archival Science* [online]. March 2001, 1(1), s. 39-55 [cit. 2017-03-31]. DOI: 10.1007/BF02435638. Dostupné z: <http://link.springer.com/10.1007/BF02435638>.
  72. DURANTI, Luciana. Diplomats: New Uses for an Old Science. *Archivaria: the Journal of the Association of Canadian Archivists* [online]. Summer 1989, (28), s. 7-27 [cit. 2017-03-16]. Dostupné z: <http://archivaria.ca/index.php/archivaria/article/view/11567/12513>.
  73. Establishing and maintaining trust in electronic records: authenticity Task Force report. [DURANTI, Luciana], ed. *The long-term preservation of authentic electronic records: the findings of the the InterPARES Project* [online]. Vancouver: The InterPARES Project, [2005 cit. 2017-03-14]. Dostupné z: [http://www.interpares.org/book/interpares\\_book\\_d\\_part1.pdf](http://www.interpares.org/book/interpares_book_d_part1.pdf).
  74. DURANTI, Luciana a Heather MACNEIL. The Protection of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project. *Archivaria: the Journal of the Association of Canadian Archivists* [online]. Fall 1996, (42), s. 46-67 [cit. 2017-03-16]. Dostupné z: <http://archivaria.ca/index.php/archivaria/article/view/12153/13158>.
  75. DURANTI, Luciana. Reliability and Authenticity: The Concepts and Their Implications. *Archivaria: the Journal of the Association of Canadian Archivists* [online]. Spring 1995, (39), s. 5-10 [cit. 2017-03-16]. Dostupné z: <http://archivaria.ca/index.php/archivaria/article/view/12063/13035>.
  76. Glossary. [DURANTI, Luciana], ed. *The long-term preservation of authentic electronic records: the findings of the the InterPARES Project* [online]. Vancouver: The InterPARES Project, [2005 cit. 2017-03-14]. Dostupné z: [http://www.interpares.org/book/interpares\\_book\\_q\\_gloss.pdf](http://www.interpares.org/book/interpares_book_q_gloss.pdf).
  77. Appendix 1: template for analysis. [DURANTI, Luciana], ed. *The long-term preservation of authentic electronic records: the findings of the the InterPARES Project* [online]. Vancouver: The InterPARES Project, [2005 cit. 2017-03-14]. Dostupné z: [http://www.interpares.org/book/interpares\\_book\\_j\\_app01.pdf](http://www.interpares.org/book/interpares_book_j_app01.pdf).



78. Trusting to time: preserving authentic records in the long term: preservation Task Force report. [DURANTI, Luciana], ed. *The long-term preservation of authentic electronic records: the findings of the the InterPARES Project* [online]. Vancouver: The InterPARES Project, [2005 cit. 2017-03-14]. Dostupné z: [http://www.interpares.org/book/interpares\\_book\\_f\\_part3.pdf](http://www.interpares.org/book/interpares_book_f_part3.pdf).
79. Appendix 2: requirements for assessing and maintaining the authenticity of electronic records. [DURANTI, Luciana], ed. *The long-term preservation of authentic electronic records: the findings of the the InterPARES Project* [online]. Vancouver: The InterPARES Project, [2005 cit. 2017-03-14]. Dostupné z: [http://www.interpares.org/book/interpares\\_book\\_k\\_app02.pdf](http://www.interpares.org/book/interpares_book_k_app02.pdf).
80. GIARETTA, David. *Advanced digital preservation*. New York: Springer, 2011, xxii, 510 s. ISBN 978-3-642-16808-6.
81. DEEGAN, Marilyn a Simon TANNER. Key issues in digital preservation. In: DEEGAN, Marilyn a Simon TANNER, ed. *Digital preservation*. London: Facet, c2006, s. 1-31. ISBN 1-85604-485-8.
82. UNESCO/PERSIST CONTENT TASK FORCE. *Guidelines for the selection of digital heritage for longterm preservation* [online]. Den Haag: UNESCO, March 2016, 19 s. [cit. 2017-03-17]. Dostupné z: [https://www.unesco.nl/sites/default/files/uploads/Comm\\_Info/persistcontentguidelinesfinal1march2016.pdf](https://www.unesco.nl/sites/default/files/uploads/Comm_Info/persistcontentguidelinesfinal1march2016.pdf).
83. Glossary. *Digital Preservation Handbook* [online]. 2nd ed. Glasgow: Digital Preservation Coalition, c2015 [cit. 2017-03-20]. Dostupné z: <http://www.dpconline.org/handbook/glossary>.
84. Metadata and documentation. *Digital Preservation Handbook* [online]. 2nd ed. Glasgow: Digital Preservation Coalition, c2015 [cit. 2017-03-20]. Dostupné z: <http://www.dpconline.org/handbook/organisational-activities/metadata-and-documentation>.
85. THE DIGITAL LIBRARY FEDERATION BENCHMARK WORKING GROUP. *Benchmark for Faithful Digital Reproductions of Monographs and Serials* [online]. Version 1. Washington (DC): Digital Library Federation, December 2002, 6 s. [cit. 2017-03-31]. Dostupné z: <http://old.diglib.org/standards/bmarkfin.pdf>.
86. BEARMAN, David a Jennifer TRANT. Authenticity of Digital Resources: Towards a Statement of Requirements in the Research Process. *D-Lib Magazine* [online]. June 1998, 4(6) [cit. 2017-03-13]. Dostupné z: <http://www.dlib.org/dlib/june98/06bearman.html>.
87. BENJAMIN, Walter. *Dílo a jeho zdroj*. Praha: Odeon, 1979, 428 s.
88. *Data Seal of Approval : Guidelines version 2017-2019* [online]. Data Seal of Approval Board, November 10, 2016, 28 s. [cit. 2017-03-20]. Dostupné z: [https://assessment.datasealofapproval.org/guidelines\\_54/pdf/](https://assessment.datasealofapproval.org/guidelines_54/pdf/).
89. ROTHENBERG, Jeff. Preserving Authentic Digital Information. In: *Authenticity in a digital environment*. Washington (DC): Council on Library and Information Resources, May 2000, s. 51-68. ISBN 1-887334-77-7.
90. GLADNEY, Henry M. *Preserving digital information*. Berlin: Springer, 2007, xxiii, 314 s. ISBN 9783540378860.
91. *Adobe RGB (1998) Color Image Encoding* [online]. Version 2005-05. San Jose (CA): Adobe Systems Incorporated, May 2005, 20 s. [cit. 2017-03-21]. Dostupné z: <https://www.adobe.com/digitalimag/pdfs/AdobeRGB1998.pdf>.
92. *Digitální restaurování českého filmového dědictví* [online]. Praha: Národní filmový archiv, 2017 [cit. 2017-03-21]. Dostupné z: <http://eea.nfa.cz/>.
93. LICHTENBERGOVÁ, Edita et al. *Katalogizace podle RDA ve formátu MARC 21- tištěné a elektronické monografie - katalogizace na úrovni minimálního/doporučeného záznamu* [online].

- Praha: Národní knihovna ČR, 2014, aktualizace srpen 2016, 154 s. [cit. 2017-03-21]. Dostupné z: <https://www.nkp.cz/o-knihovne/odborne-cinnosti/zpracovani-fondu/katalogizacni-politika/katalogizace-podle-rda-ve-formatu-marc-21-tistene-a-elektronicke-monografie-katalogizace-na-urovni-minimalniho-doporuceneho-zaznamu>.
94. Getting Started with JHOVE. *JHOVE: JSTOR/Harvard Object Validation Environment* [online]. Boston Spa: Open Preservation Foundation, c2015, updated October 2015 [cit. 2017-03-23]. Dostupné z: <http://jhove.openpreservation.org/getting-started/>.
95. HUTAŘ, Jan. Identifikace formátů – jednorázový nebo opakovaný proces? In: *CDA 2016: formátové výzvy LTP: zborník príspevkov z 1. medzinárodnej konferencie o dlhodobej archivácii*. Bratislava: Univerzitná knižnica v Bratislave, 2016, s. 35-43. ISBN 978-80-89303-51-9. ISSN 2453-9406.
96. *Kramerus* [online]. Praha: Národní knihovna ČR, c2003-2010 [cit. 2017-01-23]. Dostupné z: <http://kramerus.nkp.cz/kramerus/Welcome.do>.
97. MCLEOD, R., P. WHEATLEY a P. AYRIS. *Lifecycle information for e-literature: full report from the LIFE project* [online]. London: LIFE project, 2006 [cit. 2017-03-24]. Dostupné z: <http://discovery.ucl.ac.uk/1854/1/LifeProjMaster.pdf>.
98. FEDERAL AGENCIES DIGITIZATION INITIATIVE, Still Image Working Group. *Guidelines for TIFF Metadata: Recommended Elements and Format* [online]. Version 1.0. Washington (DC): FADGI, Feb 10, 2009, 5 s. [cit. 2017-03-15]. Dostupné z: [http://www.digitizationguidelines.gov/guidelines/TIFF\\_Metadata\\_Final.pdf](http://www.digitizationguidelines.gov/guidelines/TIFF_Metadata_Final.pdf).
99. ISO 2108:2005. *Information and documentation - International standard book number (ISBN)*. 4th ed. Geneva: ISO, 2005, 21 s.
100. ISO 3297:2007. *Information and documentation - International standard serial number (ISSN)*. 4th ed. Geneva: ISO, 2007, 20 s.
101. *Handle.Net Registry* [online]. Preston (VA, USA): Corporation for National Research Initiatives, c2015 [cit. 2017-03-30]. Dostupné z: <http://handle.net/>.
102. ISO 26324:2012. *Information and documentation - Digital object identifier system*. Geneva: ISO, 2012, 17 s.
103. HAKALA, J. RFC 3188. *Using National Bibliography Numbers as URNs* [online]. Reston: The Internet Society, Oct 2001 [cit. 2017-03-27]. Dostupné z: <http://www.ietf.org/rfc/rfc3188.txt>.
104. INTERNATIONAL ISBN AGENCY. *ISBN Users' Manual* [online]. 6th ed. London: International ISBN Agency, 2012, 31 s. [cit. 2017-03-27]. ISBN 978-92-95055-02-5. Dostupné z: <https://www.isbn-international.org/sites/default/files/ISBN%20Manual%202012%20-corr.pdf>.
105. SCHWEIZ, Bundesamt für Kultur. *E-Helvetica: URN-Handbuch* [online]. Version 1.1. Bern: Schweizerische Nationalbibliothek NB, August 2010, update Oktober 2016 [cit. 2017-03-27]. Dostupné z: <https://goo.gl/VGaChp>.
106. BELLINI, Emanuele, Chiara CIRINNA a Maurizio LUNGHI. *Persistent identifiers for cultural heritage* [online]. Glasgow: DigitalPreservationEurope (DPE), 2008 [cit. 2012-05-30]. Dostupné z: [http://www.digitalpreservationeurope.eu/publications/briefs/persistent\\_identifiers.pdf](http://www.digitalpreservationeurope.eu/publications/briefs/persistent_identifiers.pdf).
107. JEŘÁBEK, Antonín. 43. výroční zasedání Mezinárodní agentury ISBN a 23. výroční zasedání Mezinárodní agentury ISMN. *Ikaros: elektronický časopis o informační společnosti* [online].

- 2015, 19(12) [cit. 2017-03-16]. Dostupné z: <https://ikaros.cz/43-vyrocní-zasedání-mezinárodní-agentury-isbn-a-23-vyrocní-zasedání-mezinárodní-agentury-ismn>.
108. HAKALA, Juha. Document Description and Access : New Challenges. In: *CASLIN 2001: popis a zpřístupnění dokumentů : nová výzva* [online]. Praha: Národní knihovna ČR, 2001, posl. aktualizace: 28.6.2001 [cit. 2017-03-27]. Dostupné z: <http://klement.nkp.cz/caslin/caslin01/sbornik/hakala.html>.
  109. MARC 21 Format for Bibliographic Data: 015: National Bibliography Number. In: *MARC Standards* [online]. Washington (DC): Library of Congress, 09/24/2013 [cit. 2017-03-30]. Dostupné z: <https://www.loc.gov/marc/bibliographic/bd015.html>.
  110. Číslo ČNB v SK ČR. *Souborný katalog ČR - Portál CASLIN* [online]. Praha: Národní knihovna ČR, 31. 03. 2015 [cit. 2017-03-30]. Dostupné z: <http://www.caslin.cz/caslin/spoluprace/sluzby-souborneho-katalogu-cr/cislo-cnb-v-sk-cr>.
  111. PISTORIUS, Vladimír. *Jak se dělá kniha: příručka pro nakladatele*. 3., dopl. a přeprac. vyd. Příbram: Pistorius & Olšanská, 2011. ISBN 978-80-87053-50-8.
  112. ABRAMS, Stephen L. a David SEAMAN. Towards a global digital format registry. In: *World Library and Information Congress: 69th IFLA General Conference and Council "Access point library: media - information - culture": Berlin, August 1st - 9th 2003* [online]. Hague: IFLA, 2003 [cit. 2017-03-31]. Dostupné z: [https://archive.ifla.org/IV/ifla69/papers/128e-Abrams\\_Seaman.pdf](https://archive.ifla.org/IV/ifla69/papers/128e-Abrams_Seaman.pdf).
  113. BARTOŠEK, Miroslav. Archivematica – open source systém pro digitální archivaci. *Knihovna*. 2015, 26(2), s. 25-38.
  114. BRITISH LIBRARY, Digital Preservation Team. *PDF format preservation assessments* [online]. Version 1.3. London: British Library, 04/09/2015, 14 s. [cit. 2017-03-21]. Dostupné z: [http://wiki.dpconline.org/images/e/e8/PDF\\_Assessment\\_v1.3.pdf](http://wiki.dpconline.org/images/e/e8/PDF_Assessment_v1.3.pdf).
  115. FEDERAL AGENCIES DIGITIZATION INITIATIVE, Still Image Working Group. *Raster Still Images for Digitization: A Comparison of File Formats* [online]. Part 1. Detailed Matrix (unified large table). Washington (DC): FADGI, Revised, Sept 2, 2014, 3 s. [cit. 2017-03-15]. Dostupné z: [http://www.digitizationguidelines.gov/guidelines/FADGI\\_RasterFormatCompare\\_p1\\_20140902\\_r.pdf](http://www.digitizationguidelines.gov/guidelines/FADGI_RasterFormatCompare_p1_20140902_r.pdf).
  116. *The InterPARES 2 Project Glossary* [online]. Vancouver: The InterPARES Project, Current as of July 26, 2016 [cit. 2016-07-26]. Dostupné z: [http://www.interpares.org/ip2/ip2\\_term\\_pdf.cfm?pdf=glossary](http://www.interpares.org/ip2/ip2_term_pdf.cfm?pdf=glossary).
  117. HENSHAW, Christy. Simplifying our JPEG2000 conversion workflow. In: *Wellcome Library* [online]. 09/09/2011 [cit. 2017-03-13]. Dostupné z: <http://blog.wellcomelibrary.org/2011/09/simplifying-our-jpeg2000-conversion-workflow/>.
  118. HUTAŘ, Jan. *Digitalizace, popis pomocí metadat a jejich formáty* [online]. Praha, 2012, 244 s. Dizertační práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí práce Stanislav Kalkus.
  119. *JHOVE - JSTOR/Harvard Object Validation Environment* [online]. Ann Arbor (MI): JSTOR, c2003-2009, Last upd. 2009-02-25 [cit. 2016-11-02]. Dostupné z: <http://jhove.sourceforge.net/index.html>.
  120. Tutorial: using JHOVE. *JHOVE - JSTOR/Harvard Object Validation Environment* [online]. Ann Arbor (MI): JSTOR, c2004-2012, Last upd. 2012-12-05 [cit. 2016-11-02]. Dostupné z: <http://jhove.sourceforge.net/using.html>.

121. OCLC/RLG PREMIS WORKING GROUP. *Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community: A Report by the PREMIS Working Group* [online]. Dublin (OH, USA): OCLC, September 2004, 66 s. [cit. 2017-03-27]. Dostupné z:  
<http://www.oclc.org/content/dam/research/activities/pmwg/surveyreport.pdf>.
122. VYCHODIL, Bedřich. *Produkce digitálních obrazových dat a jejich kontrola*. Praha, 2013, 322 s. Dizertační práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí práce Jiří Souček.

## Seznam zkratk

AIP	archival information package
CCSDS	Consultative Committee for Space Data Systems
CD	compact disc
CDO	content data object
CRL	Center for Research Libraries
čČNB	číslo České národní bibliografie
ČIDLO	Český identifikační a lokalizační systém
ČR	Česká republika
DIP	dissemination information package
DLF	Digital Library Federation
DOI	Digital Object Identifier
DPC	Digital Preservation Coalition
DROID	Digital Record Object Identification
DVD	digital video disc
EXIF	Exchangeable image file format
FADGI	Federal Agencies Digital Guidelines Initiative
FRBR	Functional Requirements for Bibliographic Records
FTP	file transfer protocol
GDFR	Global Digital Format Registry
GIF	Graphics interchange format
HTML	HyperText Markup Language
ICC	International Color Consortium
IEC	International Electrotechnical Commission
IFLA	International Federation of Library Associations and Institutions
InterPARES	International Project on Permanent Authentic Records in Electronic Systems
iPRES	International Conference on Preservation of Digital Objects
ISBN	International Standard Book Number
ISSN	International Standard Serial Number
ISO	International Organization for Standardization
JHOVE	JSTOR/Harvard Object Validation Environment
JISC	Joint Information Systems Committee
JPEG	Joint Photographic Experts Group
LCCN	Library of Congress Control Number
LIFE	Lifecycle Information For E-Literature
LZW	Lempel–Ziv–Welch
MARC	MAchine-Readable Cataloging
MD5	Message Digest 5
METS	Metadata Encoding and Transmission Standard
MIX	Metadata for Images in XML Standard
MODS	Metadata Object Description Schema
NARA	National Archives and Records Administration
NBN	national bibliographic number

NDK	Národní digitální knihovna
NISO	National Information Standards Organization
NK ČR	Národní knihovna České republiky
OAI	Open Archives Initiative
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OAIS	Open Archival Information System
OASIS	Organization for the Advancement of Structured Information Standards
OCLC	Online Computer Library Center
PDF	Portable document format
PLANETS	Preservation and Long-term Access to Our Cultural and Scientific Heritage
PNG	Portable network graphics
PPI	pixel per inch
PREMIS	Preservation Metadata Implementation Strategies
PUID	PRONOM Persistent Unique Identifier
RLG	Research Libraries Group
SHA-1	Secure Hash Algorithm 1
SIP	submission information package
SPAR	Système de Préservation et d'Archive Réparti
SSL	Secure Sockets Layer
TIFF	Tagged Image File Format
TRAC	Trustworthy Repositories Audit & Certification
UDFR	Unified Digital Formats Registry
UK	United Kingdom
UNESCO	United Nations Educational, Scientific and Cultural Organization
ÚNMZ	Úřad pro technickou normalizaci, metrologii a státní zkušebnictví
URL	Universal Resource Locator
URN	Uniform Resource Name
USA	United States of America
XML	eXtensible Markup Language

## Seznam tabulek v textu

### Tabulky

Tabulka 1 - Požadavky pro posouzení autenticity: část 1/2 (InterPARES).....	84
Tabulka 2 - Požadavky pro posouzení autenticity: část 2/2 (InterPARES).....	85
Tabulka 3 - Požadavky pro udržování autenticity (InterPARES).....	87
Tabulka 4 - Sedm cílů digitální archivace (Caplanová) .....	95

## Evidence výpůjček

### Prohlášení:

Dávám svolení k půjčování této disertační práce. Uživatel potvrzuje svým podpisem, že bude tuto práci řádně citovat v seznamu použité literatury.

V Praze, 30. 3. 2017

Ladislav Cubr

[illegible]